

ESCUELA SUPERIOR POLITECNICA DE CHIMBORAZO



FACULTAD DE CIENCIAS

ESCUELA DE FÍSICA Y MATEMÁTICA

CARRERA DE INGENIERÍA EN ESTADÍSTICA INFORMÁTICA

**MODELO ECONOMETRICO PARA LA PRODUCCIÓN Y EL PRONÓSTICO
EN LA EMPRESA DE HILANDERÍA DE LA PARROQUIA SALINAS EN EL
CANTON GUARANDA DE LA PROVINCIA DE BOLÍVAR.**

TESIS DE GRADO

Previa a la obtención del título de:

INGENIERA EN ESTADÍSTICA INFORMÁTICA

PRESENTADO POR:

LAURA CECILIA ROCHINA CHIMBO.

RIOBAMBA - ECUADOR.

2013

AGRADECIMIENTO.

A la Escuela Superior Politécnica de Chimborazo Facultad de Ciencias,
¡Escuela de Física y Matemática.

Por darme la oportunidad de culminar mi formación académica.

A los señores doctor, Jorge Congacha. Director de tesis y al doctor, Marcelo Cortéz. Por su paciencia, por el apoyo, incondicional de realizar este trabajo investigativo.

Al personal de la Hilandería Intercomunal Salinas por permitirme realizar este trabajo investigativo

DEDICATORIA

A Dios por estar conmigo en cada momento de mi vida y darme la oportunidad de alcanzar mis metas

A mis padres Luis Rochina y Francisca Chimbo, de quienes he recibido todo el amor y el apoyo para culminar esta etapa de mi vida.

A mi esposo, mis hijos Rimaël y César por brindarme el apoyo y el cariño constantemente en mi hogar.

HOJA DE FIRMAS

Nombre	Firma	Fecha
Silvio Álvarez DECANO FACULTAD DE CIENCIAS
Dra. Jenny Orbe DIRECTORA ESCUELA DE FISICA Y MATEMATICA
Dr. Jorge Congacha DIRECTOR DE TESIS
Mat. Marcelo Cortez MIEMBRO DEL TRIBUNAL

AUTORÍA

“Yo Laura Cecilia Rochina Chimbo, soy responsable de las ideas doctrinas y el resultado expuesto en esta Tesis, y el patrimonio intelectual de la Tesis de Grado pertenece a la Escuela Superior Politécnica de Chimborazo”

.....

Laura Cecilia Rochina Chimbo
C.I.: 02156434-1

INDICE DE ABREVIATURAS

FUNORSAL	Fundación de Organizaciones Campesinas de Salinas
H.I.T	Hilandería Intercomunal Salinas
MCO	Mínimos cuadrados ordinarios
SSE	Suma de los cuadrados del error (por sus siglas en inglés sum of squared errors)
VIF	Factor de expansión de la varianza (, variance, inflactor, factor, por sus siglas en inglés)
A	Coeficiente de asimetría
CK	Coeficiente de kurtosis
CV	Coeficiente de variación
MAPE	Error porcentual absoluto medio (por sus siglas en inglés, mean absolute porcentage error)
MAD	Desviación absoluta media (por sus siglas en inglés, mean absolute desviation)
MSE	Error cuadrático medio (por sus siglas en inglés, mean squared error)

MPE	Error porcentual medio (por sus siglas en inglés, mean porcentaje error)
MA	Modelos de media móvil
AR	Modelos autorregresivos
AED	Análisis Exploratorio de Datos
ARIMA	Modelos de media móvil autorregresivo integrado

INDICE DE CONTENIDOS

AGRADECIMIENTO.....	- 2 -
DEDICATORIA.....	- 3 -
HOJA DE FIRMAS	- 4 -
AUTORÍA	- 5 -
INDICE DE ABREVIATURAS	- 6 -
INDICE DE CONTENIDOS	- 8 -
INDICE DE FIGURAS	- 12 -
INDICE DE TABLAS	- 13 -
INDICE DE GRÁFICOS	- 14 -
RESUMEN	- 15 -
SUMMARY.....	¡Error! Marcador no definido.
INTRODUCCIÓN	- 17 -
ANTECEDENTES	- 19 -
JUSTIFICACIÓN	- 20 -
OBJETIVOS	- 22 -
OBJETIVO GENERAL	- 22 -
OBJETIVOS ESPECÍFICOS	- 22 -
CAPÍTULO I	- 23 -
1. MARCO TEORICO.....	- 24 -

1.1.	Modelos económicos.....	- 24 -
1.2.	Econometría	- 26 -
1.3.	Modelos econométricos.....	- 27 -
1.3.1.	Componentes de un modelo econométrico: variables, parámetros y relaciones.....	- 28 -
1.3.2.	Utilidades de los modelos econométricos	- 29 -
1.3.3.	Metodología de la econometría.....	- 30 -
1.3.4.	Tipos de datos en econometría.....	- 31 -
1.3.5.	Clasificación de los modelos econométricos.	- 32 -
1.4.	Pronósticos.....	- 33 -
1.4.1.	Metodología de los pronósticos	- 34 -
1.4.2.	Tipos de pronósticos	- 36 -
1.5.	Componentes de series de tiempo	- 37 -
1.5.1.	Tendencia.	- 37 -
1.5.2.	Variación cíclica:	- 38 -
1.5.3.	Estacional:	- 39 -
1.5.4.	Irregular:	- 40 -
1.5.5.	Estacionariedad:	- 41 -
1.6.	Modelos que relacionan los valores observados de una serie de tiempo con sus componentes	- 42 -
1.6.1.	Modelo aditivo.....	- 42 -
1.6.2.	Modelo multiplicativo.....	- 42 -
1.6.3.	Modelo mixto.....	- 43 -
1.7.	Análisis de regresión lineal.....	- 44 -
1.7.1.	Regresión lineal simple.....	- 44 -

1.7.2.	Regresión múltiple	- 48 -
1.8.	Extensiones al modelo de regresión múltiple	- 49 -
1.8.1.	Multicolinealidad	- 49 -
1.8.2.	Heteroscedasticidad	- 53 -
1.8.3.	Autocorrelación	- 55 -
1.9.	Medición del error de pronóstico	- 59 -
1.9.1.	Notación básica de pronósticos	- 59 -
1.9.2.	Medidas de precisión	- 60 -
1.10.	Modelo autorregresivo (AR)	- 63 -
1.11.	Modelo de media móvil (MA).....	- 64 -
1.12.	Modelos de medias móviles autorregresivos (ARMA).....	- 65 -
1.13.	Modelos ARIMA	- 66 -
1.13.1.	Aplicación de una estrategia para la construcción de un modelo (ARIMA)	- 68 -
1.14.	Gruppo salinas	- 74 -
1.14.1.	Hilandería Intercomunal Salinas (H.I.S.).....	- 77 -
2.	METODOLOGÍA.....	- 79 -
2.1.	Recolección de datos	- 79 -
2.1.1.	Análisis exploratorio de datos (AED).....	- 81 -
2.1.1.1.	Etapas del AED	- 82 -
2.1.1.2.	Preparación de los datos	- 83 -
2.1.1.3.	Análisis estadístico unidimensional	- 84 -
2.1.2.	Estadísticos para el análisis exploratorio de datos (AED)	- 84 -
2.2.	MODELACION DE DATOS	- 87 -

2.2.1. Modelos de pronósticos	- 88 -
2.2.2. Exploración de patrones de datos de series de tiempo	- 88 -
2.2.3. Selección de variables significativas	- 89 -
2.2.4. Selección de un modelo significativo	- 90 -
CAPÍTULO III	- 93 -
3. Elaboración del modelo econométrico para el pronóstico de la producción de hilo.....	- 94 -
3.1. Matriz de datos	- 94 -
3.2. Análisis exploratorio de datos(AED)	- 96 -
3.2.1. Análisis de datos atípicos	- 99 -
3.3. Análisis de series de tiempo	- 104 -
3.3.1. Análisis de estacionariedad.	- 109 -
3.3.2. Forma del modelo	- 110 -
3.3.3. Elaboración del modeloEconométrico para la producción de la H.T.S.	- 112 -
3.3.4. Evaluación del modelo para el pronóstico	- 112 -
3.3.5. Utilización del modelo en los pronósticos.	- 117 -
3.4. Planteamiento del análisis estructural	- 121 -
3.4.1. Análisis de correlación.	- 121 -
3.4.2. Análisis de regresión.....	- 122 -
3.4.3. Significancia del modelo	- 126 -
3.4.4. Variables explicativas individuales:.....	- 127 -
CAPÍTULO IV	- 129 -
4. CONCLUSIONES Y RECOMENDACIONES	- 130 -
4.1. CONCLUSIONES.....	- 130 -

4.2. RECOMENDACIONES.	- 132 -
BIBLIOGRAFÍA	- 134 -
ANEXOS	- 140 -
Anexo 1	- 141 -

INDICE DE FIGURAS

Figura 1: Estrategia para la construcción de un modelo ARIMA	- 68 -
Figura 2: Estructura del GRUPPO Salinas.....	- 76 -
Figura 3: Recolección de datos.....	- 79 -
Figura 4: Diagrama de flujo de la estrategia de la construcción del modelo	- 87 -

INDICE DE TABLAS

Tabla 1: Componentes de un Serie de tiempo	- 88 -
Tabla 2: Matriz de datos.	- 94 -
Tabla 3: Estadísticos descriptivos	- 96 -
Tabla 4: Transformación de variables	- 102 -
Tabla 5: Estadísticos descriptivos de la matriz transformada	- 103 -
Tabla 6: Estadísticos para la evaluación del modelo de pronósticos.....	- 113 -
Tabla 7: Comparación de metodologías de Winters.....	- 117 -
Tabla 8: Valores mediante el método de Winters para la producción de hilo ..	- 118 -
Tabla9: Pronósticos de la producción de hilo	- 120 -
Tabla 10: Correlación de las variables endógena y exógenas	- 121 -
Tabla 11: Tabla ANOVA	- 123 -
Tabla 12: Resultados por variables exógenas.....	- 124 -
Tabla 13: ANOVA	- 126 -
Tabla 14: Tabla de estadísticos para la significancia del modelo.	- 128 -

INDICE DE GRÁFICOS

Gráfica 1: Serie de datos con tendencia	- 38 -
Gráfica 2: Serie de datos con ciclicidad	- 39 -
Gráfica 3: Serie datos estacionales	- 40 -
Gráfica 4: Serie de datos con irregularidad	- 40 -
Gráfica 5: Serie de datos estacionarios	- 41 -
Gráfica 6: Correlograma de una serie de tiempo	- 69 -
Gráfica 7: Box-plot de la variable endógena	- 100 -
Gráfica 8: Box-plot de las variables exógenas	- 101 -
Gráfica 9: Representación de series temporales de la producción de hilo	- 105 -
Gráfica 10: Función de autocorrelación para la variable endógena (producción)	- 106 -
Gráfica 11: Representación del análisis de residuos de la variable producción	- 107 -
Gráfica 12: Diferencia de la variable endógena	- 109 -
Gráfica 13: Representación gráfica de Autocorrelación	- 110 -
Gráfica 14: Representación gráfica de Autocorrelación parcial.	- 111 -
Gráfica 15: Representación gráfica de residuos para el modelo de producción.	- 114 -
Gráfica 16: Resultados gráficos Método de Winters	- 119 -

RESUMEN

Se ha elaborado un modelo econométrico para el pronóstico de la producción de hilo en la Hilandería Intercomunal Salinas perteneciente al Cantón Guaranda en la Provincia de Bolívar, tomando en cuenta la necesidad de estudios estadístico para el buen funcionamiento de la empresa y la toma de decisiones, ya que no se han realizado estudios similares desde la creación en la industria.

El análisis se realizó en la hoja electrónica Excel y en el software estadístico Minitab mediante la recopilación de los datos de los archivo de la empresa, entonces se realizó la aplicación de técnicas como cálculo de estadística descriptiva, transformación de datos, correlación y regresión lineal; consiguiendo resultados que permiten medir la situación actual de la empresa. Se identificó un modelo de pronósticos por el método de Winters ya que la serie de datos de la producción de hilo presenta tendencia moderada y estacionalidad con variaciones, además se hizo un análisis estructural de las variables que implican la producción de hilo a través de un modelo regresión múltiple.

En conclusión durante los últimos 4 años la producción de hilo ha sido de 6.241 libras mensuales que representa el 52% de la producción supuesta de 12.000 libras, así mismo las pérdidas de materia prima (lana de llama y oveja) son de 20 quintales. Para una mejor producción y el mejoramiento de la empresa se recomienda estudios de control de calidad y realizar un mayor control en los diferentes procesos de producción para evitar las pérdidas de materia prima.

SUMMARY

This research involves a study based in an econometric model for yarn production in Hilandería Intercomunal Salinas spinning mill belonging to Guaranda canton in in the Bolívar province. It has been taken into account the necessity of a statistical study for the correct functioning of the company and decision-making as there have not been carried out similar studies since the creation of this industry.

The data analysis was made by using the Excel spreadsheet and the statistic software Minitab; after having gathered the information from the company file then, some techniques like descriptive statistic calculation, data transformation, linear correlation and regression analysis were followed, which made possible to measure the actual situation of the enterprise. It was also identified a prognostics model by the winter method as the data serial of yarn production shows a moderate tendency and stationarity with variations furthermore, it was developed a structural analysis of the variables implying yarn production through a multiple regression model. It is concluded that yarn production during the last four years has corresponded to 6241 pounds per month, which represents 52% of the production supposed to reach 12.000 pounds as well as raw material loss (llama and sheep wool) border 20 quintals. It is recommended to make quality control studies to get a better production of the company as well as its improvement and another aspect to consider is to control the several production processes in order to avoid raw material loss.

INTRODUCCIÓN

La presente investigación de grado, titulado “MODELO ECONOMETRICO PARA LA PRODUCCIÓN Y EL PRONÓSTICO EN LA EMPRESA DE HILANDERÍA DE LA PARROQUIA SALINAS EN EL CANTON GUARANDA DE LA PROVINCIA BOLÍVAR” realizado en la H.T.S. (Hilandería Intercomunal Salinas) gracias al interés de la FUNORSAL (Fundación de Organizaciones Campesinas de Salinas) organización al cual pertenece la empresa, dedicada a la producción de hilo, derivado de la lana de oveja, llama y alpaca.

La investigación se realizó con los propósitos de plantear estrategias de planificación en la producción de hilo y buscar el mejoramiento de la empresa, ya que sus utilidades van dirigidas a la ayuda social y el desarrollo de la localidad. Con el fin de, elaborar un modelo econométrico para el pronóstico de la producción de hilo por el periodo de un año, realizar un AED de las variables y determinar la significancia del modelo propuesto.

Para el desarrollo de los objetivos propuestos se ha realizado un análisis de series temporales, ya que los datos son registrados mensualmente, además se ha empleado, el método de Winters y se ha aplicado el análisis de regresión múltiple para realizar un análisis estructural.

La tesis se encuentra dividida en cuatro capítulos: el Capítulo I (Marco teórico) trata el fundamento teórico a aplicarse en el estudio. Capítulo II (Metodología) en esta parte se explica a través de diagramas de flujo las etapas como se desarrolla la tesis. Capítulo III (Elaboración del modelo econométrico para el pronóstico de la producción de hilo) es el desarrollo analítico del proyecto de tesis, tomando en cuenta cada una de las variables en estudio y realizando un análisis estructural con los mismos. Capítulo IV (Conclusiones y Recomendaciones) habiendo realizado la investigación se llegó a varias conclusiones y a citar algunas recomendaciones.

ANTECEDENTES

El desarrollo de una empresa, requiere de una eficiente planificación constante en base a indicadores estadísticos que permita alcanzar los objetivos planteados en forma eficaz y con un correcto empleo de recursos asignados.

Desde 1970, la parroquia Salinas empezó a enfrentar la pobreza y marginación, con el apoyo de voluntarios extranjeros y la Misión Salesiana, con la creación de varias industrias entre ellas la Hilandería, donde la materia prima es la lana de llama y oveja, estos son sometidos a diversos tipos de procesos como la recepción de la materia prima, secada, lavado y nuevamente secado para llevarla a la maquina encargada de sacar las fibras e hilarlo.

Anteriormente en la empresa no se ha realizado estudios estadísticos, lo que hace que sea muy relevante el trabajo ejecutado sin desconocer el esfuerzo de los administradores y voluntarios extranjeros encargados de la Hilandería.

JUSTIFICACIÓN

Debido a la necesidad de conocer la evolución de la producción de hilo presentada por la H.T.S. en los últimos 3 años y saber un valor estimado en meses posteriores, se plantea la presente investigación de tesis, ya que el desconocimiento de técnicas estadísticas y el personal capacitado en esta área no ha permitido realizar estudios similares desde los inicios de funcionamiento de la empresa.

Esta investigación nos ayuda a evaluar administraciones anteriores, planificar estrategias, actividades, además desarrollar políticas para el cumplimiento de la demanda, mejoramiento y optimización de la producción, la información que se obtiene es confiable ya que se recogió datos de los registros originales que actualmente se encuentran archivados digital e impresos a cargo del director ejecutivo de la FUNORSAL, y en la planta administrativa de la empresa, los mismos que fueron registrados en cada proceso de la planta de producción.

También son analizados con una base teórica sustentable y con las debidas comprobaciones; lo que arrojó resultados muy útiles para tomar decisiones de políticas, actividades, estrategias de mercado, productos etc, e incentivar la producción de los animales bovinos ya que allí podemos encontrar la materia prima y fundamental para el funcionamiento de la empresa, los mismos que generan empleo de forma directa e indirecta, así poder contribuir con el desarrollo

social y económico, para disminuir el índice de analfabetismo y desempleo, de las comunidades que contribuyen al funcionamiento de la hilandería. Tomando en cuenta, las utilidades que genera la empresa son destinadas a la ayuda social de la localidad.

OBJETIVOS

OBJETIVO GENERAL

Elaborar un modelo econométrico para la producción y pronósticos de la empresa de Hilandería.

OBJETIVOS ESPECÍFICOS

- Realizar un análisis exploratorio de datos de la producción hilo y las variables que implica el mismo, de los últimos años para la construcción del modelo.
- Determinar la significancia de las variables definidas en el modelo.
- Pronosticar la producción mensual de hilo por el periodo de un año

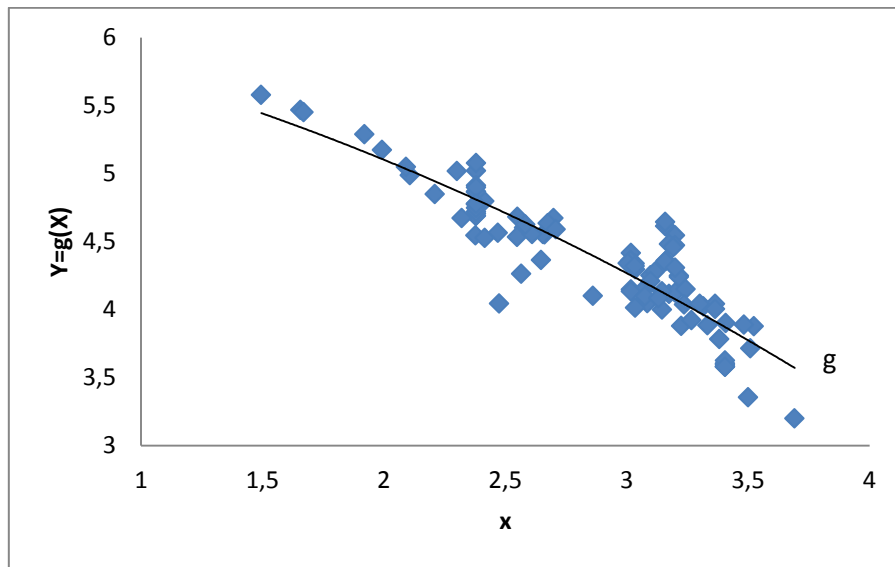
CAPÍTULO I

1. MARCO TEORICO

1.1. Modelos económicos.

Un modelo es una representación simplificada de la realidad que recoge los aspectos fundamentales de la misma que tienen interés para los objetivos del investigador o analista. Los modelos económicos plantean normalmente la relación entre variables económicas. Así la relación entre el precio y la cantidad demandada de un bien, la dependencia entre la tasa de inflación y el nivel de desempleo o los salarios, son ejemplos de la relación entre variables que los modelos económicos están interesados.

En la teoría económica, estas relaciones están caracterizadas por funciones matemáticas exactas. Es decir, el economista postula que la cantidad demandada de un bien es una función decreciente del precio del mismo. Por tanto si se denomina a Y por la cantidad demandada, y X por el precio, podemos postular una relación como:



$$Y=g(X)$$

Donde $g(\cdot)$ es una función en sentido matemático, es decir, a cada valor del precio le corresponderá un valor de la cantidad demandada. A partir de este planteamiento, al economista le preocupa responder ¿cuál será la forma que adopta $g(\cdot)$? ¿será $g(X)$ creciente o decreciente con X ? Sin embargo, cuando pasamos a las relaciones empíricas observamos que en el mundo real las variables Y y X adoptan unos valores concretos, y será dicha combinación de pares de valores la que determine la forma de $g(\cdot)$.

Los modelos económicos se caracterizan por construcciones teóricas para reflejar sistemas y procesos económicos. Generalmente ello se hace definiendo un conjunto de variables y un conjunto de relaciones definidas entre los entes representados por las variables. Algunas de ellas pueden tomar

valores constantes específicos según la característica del modelo, se convierten realmente en parámetros en muchos casos, dependiendo de lo que se intente al efectuar la modelación.

1.2. Econometría

La econometría es una ciencia que surge por la necesidad de cuantificar los modelos económicos para tratar de contrastar si las diversas teorías económicas se cumplen cuando se ven enfrentadas a los datos, la importancia de los mismos, debe señalarse que, la necesidad de contrastar si las teorías económicas se adecuan a los datos observados o la convivencia de desarrollar modelos ajustados a la realidad si los más simples son rechazados por los datos a los que se enfrentan, hace que se requieran bien nuevas teorías u otro tipo de información para el contraste de algunos modelos que no son susceptibles de ser contrastados con los datos disponibles.

Es importante señalar que la econometría nos proporciona una metodología para alcanzar un fin. Su uso adecuado proporciona una poderosa herramienta para el análisis de fenómenos económicos u otras áreas

1.3. Modelos econométricos.

Modelo econométrico es un modelo económico que contiene las especificaciones necesarias para su aplicación empírica, viene expresado así:

$$Y = g(X) + u$$

Donde u es la parte aleatoria no explicada por X . Estos elementos aleatorios, en ocasiones llamados errores, recogen la diferencia entre el valor observado de Y y el valor que mediante $g(X)$ la recta de regresión asignamos, y pueden deberse a errores de medición en la variable o a imperfecciones en la especificación de $g(X)$, entre otros. El conjunto de especificaciones que requiere un modelo econométrico son:

- **Identificar las variables** que fundamentalmente influyen sobre el aspecto que se desea estudiar
- **Formular una relación** o forma funcional concreta entre el conjunto de variables (aquella que se desea explicar y las consideradas como influyentes en ella).

- **Introducir un término denominado “perturbación aleatoria”** lo que permite razonar en términos probabilísticos y no exactos.

1.3.1. Componentes de un modelo econométrico: variables, parámetros y relaciones.

Variables: las variables son los factores o entes elementales que actúan en un fenómeno desde el punto de vista cuantitativo”. En las matemáticas, las variables se dividen en:

- variables dependientes
- variables independientes

Ahora bien, a la hora de analizar la realidad económica no resulta fácil realizar la distinción enunciada ya que son frecuentes las interrelaciones entre las variables económicas y se hace necesario, por tanto, acudir a otro tipo de clasificación.

Consecuentemente, en econometría se distingue entre:

Variables endógenas: aquellas que vienen explicadas por el funcionamiento del modelo.

Variables exógenas: son aquellas cuyos valores inciden sobre el modelo desde el exterior; es decir, son determinadas fuera del modelo pero influyen en el comportamiento de las endógenas.

Por otra parte, las variables que aparecen en un modelo se pueden referir al mismo periodo (o al mismo instante temporal) o a periodos distintos. Más concretamente, las variables de un modelo pueden referirse exclusivamente a un periodo t , o a los periodos $t-2$, $t-1$, t , ..., .En este último caso decimos que el modelo contiene variables retardadas en el tiempo.

1.3.2. Utilidades de los modelos econométricos

El modelo econométrico tiene tres utilidades principales:

Análisis estructural: cuantificación de las relaciones que entre el periodo analizado ha existido entre las variables implicadas, a través del conocimiento del signo y valor de los parámetros estimados. Es decir, sirve para conocer como inciden en las endógenas variaciones de las variables exógenas.

Predicción: Dados valores a futuro para las variables exógenas, y conociendo la expresión matemática que relaciona las variables exógenas y endógena, es posible predecir los valores que tomará a futuro la variable objeto de estudio.

Simulación o evaluación de políticas: Efectos que tienen sobre la endógena diferentes estrategias que se planteen de las variables exógenas. Por ejemplo si analizamos las ventas de una empresa en función de los precios del producto y del nivel de gasto realizado en publicidad, podríamos estar interesados en analizar cuánto incrementarían las unidades vendidas si se mantienen los precios fijos y se incrementa el gasto en publicidad en un porcentaje determinado.

1.3.3. Metodología de la econometría

La Econometría trata de realizar inferencias sobre datos que no son susceptibles de repetición, en unos casos para explicar relaciones causales y en otros para tratar de realizar predicciones para las variables de interés.

Aunque existen diversas escuelas de pensamiento sobre metodología econométrica, a continuación se presenta la metodología clásica, que predomina en la investigación empírica en economía y en los campos

relacionados. En términos generales, la metodología de la econometría tradicional se realiza dentro de los siguientes lineamientos:

- **Especificación:** Esta etapa comprende tanto la determinación del tema objeto de análisis como la definición de las variables explicativas que se incluirán en el modelo.
- **Estimación:** consiste en el cálculo del valor de los parámetros Para realizar esta fase es necesario previamente haber realizado una búsqueda y depuración de datos. Es necesaria la obtención de datos suficientes, homogéneos y actualizados.
- **Validación:** a través de la interpretación de los resultados analizaremos la bondad del modelo.

1.3.4. Tipos de datos en econometría

Datos de series de tiempo: los datos pueden corresponder a los valores de una variable en el tiempo. Estos pueden tener frecuencia por ejemplo: diaria, semanal, mensual o anual, etc.

Datos de corte transversal: los valores corresponden a distintos sujetos para un mismo momento del tiempo. Por ejemplo de series del tipo de consumo de diferentes familias, inversión de distintas empresas, paro en diferentes provincias, etc.

Datos de panel: la combinación de datos de series temporales y corte transversal.

1.3.5. Clasificación de los modelos econométricos.

Los modelos se pueden clasificar en las siguientes categorías:

- Si solamente estamos interesados en una relación de comportamiento, por ejemplo, examinar la influencia de la inflación sobre el nivel de desempleo, los modelos serán **uniecuacionales**. En el caso en que el interés radique en estudiar las relaciones entre varias variables dependientes en conjunto de variables independientes, los modelos pueden ser **multiecuacionales**.
- Atendiendo a la forma de las relaciones los modelos pueden clasificarse en: **modelos lineales** donde todas las relaciones son lineales, **modelos no lineales** cuando alguna relación es no lineal. Esta clasificación tiene

interés por cuanto la posibilidad de la aplicación práctica de un modelo depende esencialmente de la forma de las relaciones.

- Según el momento del tiempo al que están referidas sus variables se pueden clasificar en **estáticos** o **dinámicos**. En el primer caso, todas sus variables están referidas al mismo instante del tiempo. En el segundo, figurará alguna variable en un instante de tiempo distinto.
- Los modelos **cualitativos** medidos a través de datos cualitativos también participan en los modelos econométricos. Dichas variables cualitativas pueden tener un doble rol en el modelo econométrico bien como variable endógena o como variable exógena

1.4. Pronósticos

Pronóstico es un método mediante el cual se intenta conocer el comportamiento futuro de alguna variable con algún grado de certeza. Los pronósticos son una de las herramientas fundamentales para la toma de decisiones dentro de las organizaciones tanto productivas como sin fines de lucro.

Entonces el objetivo principal de los pronósticos es reducir la incertidumbre acerca de lo que puede acontecer en el futuro proporcionando información cercana a la realidad que permita tomar decisiones sobre los cursos de acción a tomar tanto en el presente como en el futuro.

1.4.1. Metodología de los pronósticos

La consideración primordial en la elección de un método de pronóstico es que los resultados deben facilitar el proceso de la toma de decisiones de los administradores de la organización.

La metodología que produzca los pronósticos más precisos en un caso, quizás no sea la mejor en otra situación. Sin embargo el(los) método(s) elegido(s) debe(n) producir un pronóstico adecuado, oportuno y entendible para los administradores de tal forma que puedan ayudar a tomar mejores decisiones.

El reconocimiento de que las técnicas de pronósticos operan con los datos generados por sucesos históricos lleva a la identificación de los siguientes cinco pasos en el proceso de pronóstico:

1. Formulación del problema y recolección de datos
2. Depuración de datos

3. Construcción y evaluación del modelo
4. Aplicación del modelo
5. Evaluación del pronóstico

En el paso 1, formulación del problema y recolección de datos se tratan como un paso único, debido a que están íntimamente relacionados. El problema determina los datos adecuados.

El paso 2, manipulación y depuración de datos, con frecuencia es necesario. En el proceso de pronóstico es posible tener demasiados o muy pocos datos. Algunos datos pueden ser irrelevantes al problema. Otros podrían tener valores omitidos que deberían estimarse. Algunos datos podrían tener que expresarse en unidades diferentes a las originales. Otros quizá deban procesarse previamente. Otros datos podrían ser adecuados, aunque solamente ciertos periodos históricos. Por lo general, se requiere cierto esfuerzo para obtener los datos de la forma requerida con el propósito de utilizar determinados procedimientos de pronósticos.

El paso 3, construcción y evaluación del modelo, implica adecuar los datos recolectados en un modelo de pronósticos que sea adecuado en términos de minimización del error de pronóstico

El paso 4, aplicación del modelo, consiste en los pronósticos reales del modelo que se generan una vez que se han recolectado y quizás reducido a sólo los datos adecuados, tan pronto se ha elegido un modelo de pronósticos.

El paso 5, evaluación del pronóstico, implica comparar los valores del pronóstico con los valores históricos reales. En este proceso, algunos de los más recientes valores de datos se retienen del grupo de datos que se analiza.

1.4.2. Tipos de pronósticos

Los pronósticos de largo plazo son necesarios para establecer el curso general de una organización y son el enfoque exclusivo de la alta dirección.

Los pronósticos de corto plazo se usan para diseñar estrategias inmediatas, y los mandos medios y las gerencias de primera línea lo usan para cubrir las necesidades del futuro inmediato.

Los pronósticos en términos de su posición dentro de un continuo **micro y macro**; es decir, en la medida que involucren pequeños detalles en comparación con una gran escala. También pueden clasificarse según sean **cuantitativos o cualitativos**.

En un extremo, una técnica totalmente cualitativas no requiere manipulación abierta de datos solamente se utiliza el juicio de quien pronostica. Incluso aquí en realidad, el juicio de esta persona es resultado de la manipulación mental de datos históricos. En el otro extremo, las técnicas puramente cuantitativas no necesitan elementos de juicio; son procedimientos mecánicos que producen resultados cuantitativos. Por supuesto algunos procedimientos cuantitativos requieren una manipulación más sofisticada de los datos que otros.

1.5. Componentes de series de tiempo

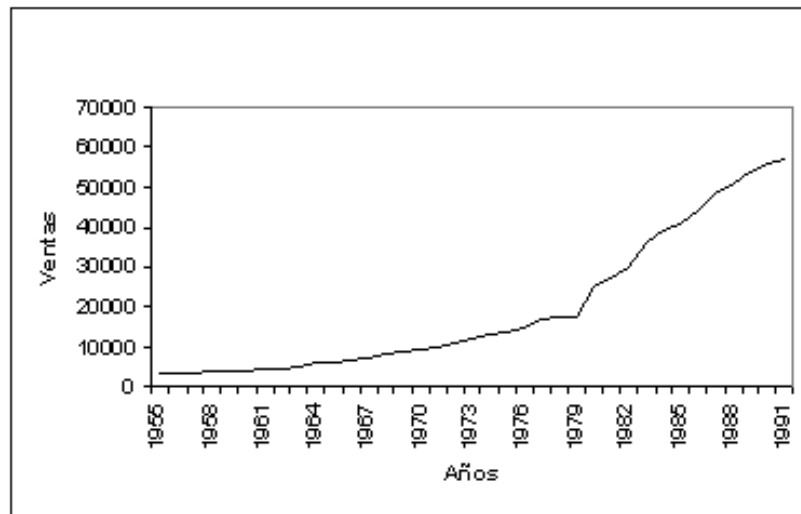
Una serie de tiempo es un conjunto de observaciones producidas en determinados momentos, generalmente a intervalos iguales. Se ha revelado que existen ciertos movimientos o variaciones que pueden medirse y observarse por separado. Estos movimientos, llamados a menudo componentes, de una serie de tiempo y que se supone son causados por fenómenos distintos son los siguientes:

1.5.1. Tendencia.

Es la componente que representa el crecimiento subyacente o la declinación de una serie de tiempo, como se presenta en la figura. Las fuerzas básicas que

producen o afectan la tendencia de una serie son: cambios en la población, inflación, cambio tecnológico e incremento en la productividad etc. La tendencia se denota con T.

Gráfica 1: Serie de datos con tendencia



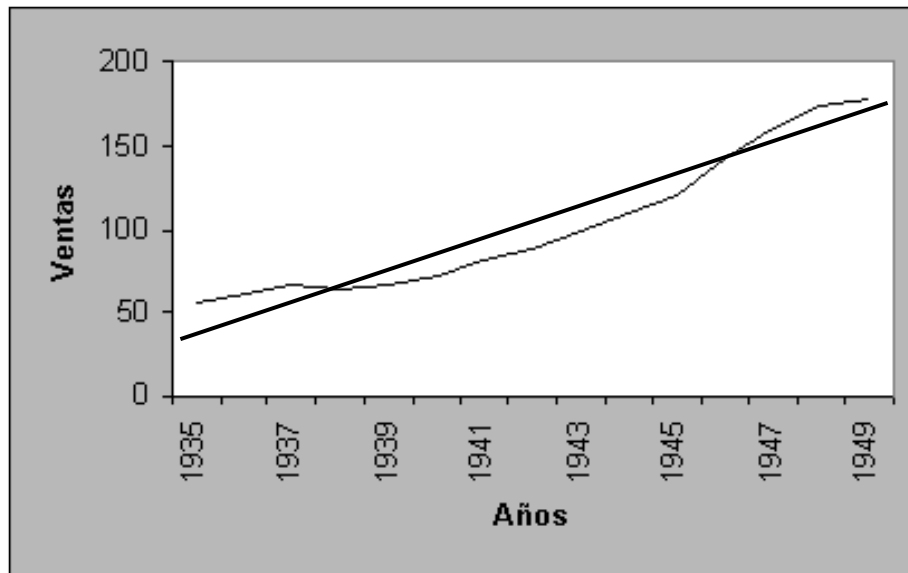
1.5.2. Variación cíclica:

El componente cíclico es una serie de fluctuaciones (como un oleaje) o ciclos de más de un año de duración. Por lo general, las condiciones económicas cambiantes, producen ciclos.

Resulta difícil identificar los ciclos y con frecuencia se le considera parte de la tendencia. En este caso, el componente del crecimiento general subyacente o

declinación es conocido como ciclo-tendencia y se denota también por la T. Se usa la notación para la tendencia porque frecuentemente el componente cíclico no puede ser separado de ésta. La **C** denota al componente cíclico.

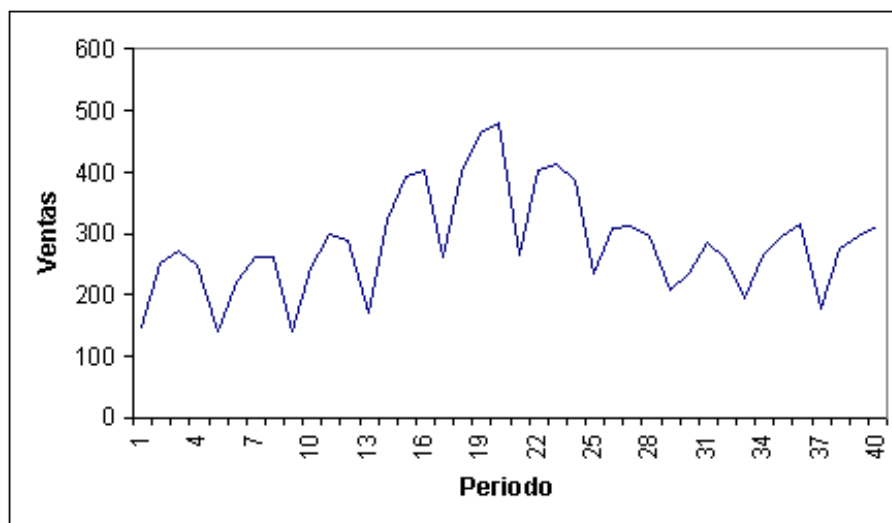
Gráfica 2: Serie de datos con ciclicidad



1.5.3. Estacional:

De manera típica las fluctuaciones estacionales se encuentran en datos trimestrales, mensuales o semanales. La variación estacional se refiere a un patrón de cambio más o menos estable que aparece anualmente y se repite con esta periodicidad. Los patrones estacionales se deben a la influencia del clima o sucesos relativos al calendario. La **S** denota al componente estacional.

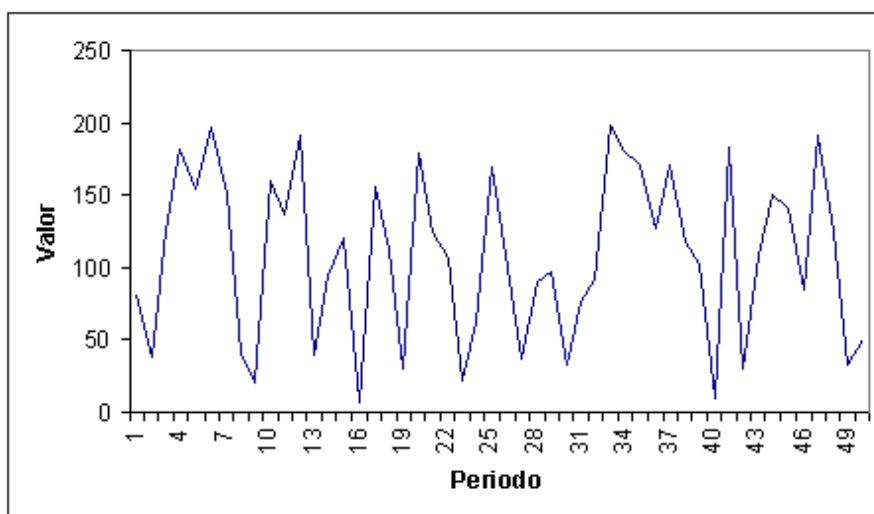
Gráfica 3: Serie datos estacionales



1.5.4. Irregular:

El componente irregular consiste en las fluctuaciones impredecibles o aleatorias. La I denota al componente irregular.

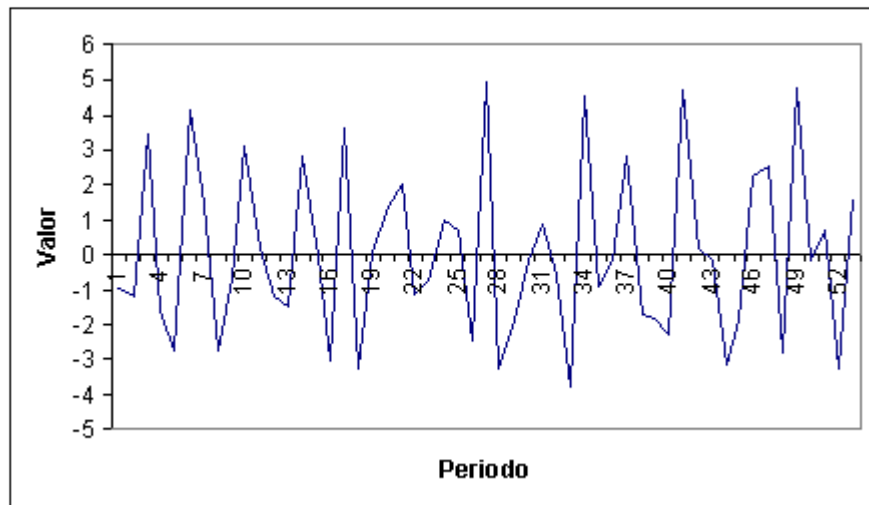
Gráfica 4: Serie de datos con irregularidad



1.5.5. Estacionariedad:

Una serie de tiempo estacionaria es aquella cuyas propiedades estadísticas básicas, tales como la media y la varianza, permanecen constantes a lo largo del tiempo. Es decir una serie que varía alrededor de un nivel fijo a lo largo del tiempo (sin crecimiento ni decrecimiento). La **ST** denota al componente estacionario.

Gráfica 5: Serie de datos estacionarios



1.6. Modelos que relacionan los valores observados de una serie de tiempo con sus componentes

1.6.1. Modelo aditivo

Un modelo que trata a los valores de la series de tiempo como la suma de sus componentes. Funciona mejor cuando la serie de tiempo que se analiza tiene aproximadamente la misma variabilidad a lo largo de la serie.

$$Y_t = T_t + S_t + I_t$$

Donde,

Y_t = Valores observados de una serie de tiempo t

T_t = Componente de tendencia en el tiempo t

S_t = Componente estacional en el tiempo t

I_t = Componente irregular en el tiempo t

1.6.2. Modelo multiplicativo

Un modelo que trata a los valores de la serie de tiempo como el producto de sus componentes. Este modelo funciona mejor cuando la variabilidad de la serie de tiempo se incrementa al aumentar el nivel.

$$Y_t = T_t \times S_t \times I_t$$

Donde,

Y_t = Valores observados de una serie de tiempo t

T_t = Componente de tendencia en el tiempo

S_t = Componente estacional en el tiempo t

I_t = Componente irregular en el tiempo t

1.6.3. Modelo mixto

Existen variantes de los modelos antes mencionados que contienen términos aditivos como multiplicativos.

$$Y_t = T_t \times S_t + I_t$$

Nota.- pueden existir otros modelos

Donde,

Y_t = Valores observados de una serie de tiempo t

T_t = Componente de tendencia en el tiempo t

S_t = Componente estacional en el tiempo t

I_t = Componente irregular en el tiempo

1.7. Análisis de regresión lineal

1.7.1. Regresión lineal simple

El análisis de la regresión simple tiene por objeto estimar la relación funcional entre dos variables. En este tipo de análisis se supone que una de las variables (Y) viene explicada por otra (X), de tal forma que se puede decir que la variable (Y) es una variable dependiente o endógena y que la variable X es la variable independiente o exógena. Viene dada por la ecuación:

$$Y = \beta_0 + \beta_1 X$$

Donde β_0 y β_1 representan, el intercepto y la pendiente de la recta de regresión respectivamente.

Las relaciones económicas entre dos variables no obedecen al comportamiento de un modelo determinístico puesto que numerosos factores, por ejemplo de la conducta de los agentes, no son observables y obedecen en numerosas ocasiones a comportamientos aleatorios.

1.7.1.1. Métodos de estimación de parámetros

Existen diversos métodos para estimar los parámetros del modelo, muchos de los cuales se basan en los residuos o errores, que se definen a continuación:

$$u_i = Y_i - \hat{Y}_i \quad i=1,2,\dots,n$$

Dónde:

u_i = Residuales

Y_i = valor observado de Y

\hat{Y}_i = valor pronosticado en Y

Mínimos cuadrados ordinarios(MCO):se utiliza para calcular la ecuación de una línea recta que minimiza la suma de las distancias al cuadrado entre los puntos de datos X-Y y la línea, medida en la dirección del eje vertical (Y). El método de mínimos cuadrados elige los valores para el intercepto β_0 y la pendiente β_1 con el eje Y, a fin de minimizar la suma de los cuadrados del error –distancias- entre los valores Y y la recta:

$$SSE = \sum (Y - \hat{Y})^2 = \sum (Y - b_0 - b_1 X)^2$$

Dónde:

SSE= suma de los cuadrados del error (por sus siglas en inglés sum of squared errors)

b_0 =intercepto estimado

b_1 =pendiente estimada

De manera particular,

$$b_1 = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2} = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$$
$$b_0 = \bar{Y} - b_1 \bar{X} = \frac{\sum Y}{n} - \frac{b_1 \sum X}{n}$$

Propiedades del estimador de mínimos cuadrados ordinarios (MCO)

- **Linealidad:** la forma funcional que liga al verdadero valor del parámetro

$$E(Y) = Y - \mu_Y$$

- **Insiesgadez:** el valor más probable del estimador coincide con el verdadero valor del parámetro

$$E(\hat{\beta}_1) = \beta_1$$

- **Eficiencia:** la desviación entre el verdadero valor del parámetro estimado y el valor del estimador será la menor posible.

$$Var(\hat{\beta}) < Var(\hat{\beta}_1)$$

- **Consistencia:** la diferencia entre el valor estimado del parámetro y el real se anula para una muestra infinita.

$$\lim_{n \rightarrow \infty} \hat{\beta}_1 = \beta_1$$

1.7.1.2. Limitaciones del modelo de regresión simple

En modelos de la teoría económica, suele haber algunas variables que afecten a la variable de interés. Es decir, existen diferencias múltiples sobre la variable de interés. El modelo de regresión simple no es capaz de captar todo este tipo de relaciones.

En ocasiones, las relaciones entre dos variables permiten predecir con precisión la variable endógena mediante el conocimiento de la exógena. Desafortunadamente, muchas situaciones de predicción no son tan simples en la práctica. Es habitual que se necesite más de una variable exógena para

predecir con precisión la variable endógena. Los modelos de regresión con más de una variable exógena se llaman modelos de regresión múltiple.

1.7.2. Regresión múltiple

La mayoría de los conceptos presentados en la regresión lineal simple se mantienen en la regresión múltiple. Sin embargo, también surgen algunos conceptos nuevos debido a que se usan más de una variable exógena para predecir la variable endógena.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K + \varepsilon$$

Donde:

Y es la variable endógena

X_1, X_2, \dots, X_K Son las variables exógenas.

ε , es el componente del error que representa las desviaciones de la respuesta respecto a la relación verdadera. Son variables aleatorias no observables que explican los efectos de otros factores en la respuesta. Se suponen que los

errores son independientes y que cada uno se distribuye normalmente, $\varepsilon \sim N(0, \sigma)$.

1.8. Extensiones al modelo de regresión múltiple

1.8.1. Multicolinealidad

La multicolinealidad hace referencia a la relación lineal entre las variables exógenas. Sus efectos sobre la estimación son inciertos dependiendo del grado de colinealidad de las variables exógenas y de la importancia que estas tengan sobre la variable endógena.

La presencia de multicolinealidad no constituye un problema, el problema es el grado con que las variables están correlacionadas, y por ende dificulta la interpretación de una variable exógena individual en la respuesta (endógena).

1.8.1.1. Detección de multicolinealidad

Para detectar la multicolinealidad no hacemos pruebas, lo que tenemos son unos indicadores sobre la existencia de multicolinealidad. Algunos de estos indicadores son:

- Elevadas correlaciones entre las variables exógenas nos indica un problema de multicolinealidad.
- El factor de inflación de la varianza: La fuerza de la multicolinealidad se mide por el factor de expansión de la varianza (VIF, variance, inflator, factor, por sus siglas en inglés)

$$VIF_j = \frac{1}{1 - R_j^2} \quad j = 1, 2, \dots, k$$

Donde,

R_j^2 Coeficiente de determinación

- ✓ Un VIF que se acerque a 1 indica que la multicolinealidad no es un problema.
- ✓ Un VIF mucho mayor que 1 indica que el coeficiente estimado sujeto a la variable exógena es inestable.
- ✓ Un VIF grande indica que existe información redundante entre las variables exógenas.

1.8.1.2. Soluciones a la multicolinealidad:

Cuando se presenta multicolinealidad, se acude a diferentes soluciones propuestas, aunque ninguna de ellas suele resultar plenamente satisfactoria. Las medidas a tomar pueden concretarse en los puntos siguientes:

1. Eliminar variables

En algunas ocasiones podemos estar interesados en unos parámetros más que en otros y, si existe un elevado grado de multicolinealidad entre las variables que forman parte de la regresión, podemos tratar de excluir una de ellas y estimar los parámetros correspondientes al resto. Sin embargo, puede generar un error de especificación del modelo. No se debe eliminar una variable de un modelo econométrico viable, sólo porque el problema de colinealidad sea grave.

2. Aumentar el tamaño muestral

Esta es una solución que a menudo se suele sugerir pero que, en numerosas ocasiones, no es factible (porque para la estimación del modelo seguramente se habrá considerado todas las observaciones disponibles para las variables) y que en otras no conduce a una eliminación o reducción del problema, porque lo relevante no es el número de observaciones, sino el contenido informativo de

las mismas, si al añadir nuevas observaciones, en las mismas se sigue cumpliendo el patrón de multicolinealidad.

- Si estamos realizando una regresión con datos anuales y se tiene un problema de multicolinealidad, podemos trimestralizar la serie o exprésalo en datos mensuales.
- Obtención de una nueva muestra o ampliar la muestra obteniendo datos adicionales, puede reducir la gravedad de la multicolinealidad.

3. Transformación de los datos

La transformación de las variables incluidas en el modelo puede minimizar, incluso resolver, el problema de la colinealidad.

- Cuando se tienen distribuciones de frecuencias con asimetría negativa (frecuencias altas hacia el lado derecho de la distribución), es conveniente aplicar la transformación $y = x^2$. Esta transformación comprime la escala para valores pequeños y la expande para valores altos.

- Para distribuciones asimétricas positivas se usan las transformaciones \sqrt{x} , $\ln(x)$ y $1/x$, que comprimen los valores altos y expanden los pequeños.

El efecto de estas transformaciones está en orden creciente: menos efecto \sqrt{x} , más $\ln(x)$ y más aún $1/x$. Se debe tener presente la teoría económica a la hora de efectuar transformaciones que tengan sentido y que no suponga meramente un juego matemático con las ecuaciones del modelo econométrico.

1.8.2. Heteroscedasticidad

Existe heteroscedasticidad cuando la varianza de los residuos o perturbaciones no es constante para diferentes observaciones. En la mayoría de los casos, esta varianza depende o es función de alguna de las variables exógenas, y dicha función puede ser creciente o decreciente.

1.8.2.1. Detección de heteroscedasticidad

- Cuando se presenta heteroscedasticidad, los estimadores MCO serán insesgados pero no eficientes, es decir, su varianza no será mínima. Al verse incrementada la varianza de los residuos, la precisión de los

coeficientes de regresión se verá reducida. Asimismo, el procedimiento de MCO no proporciona estimaciones correctas de los errores estándar de los coeficientes de regresión, ya que se han calculado bajo el supuesto de que la varianza de las perturbaciones es constante, cuando en realidad no lo es.

- De la inspección de un gráfico de residuos se puede obtener una idea clara sobre la presencia de una conducta sistemática de los residuos. Para el caso en que la varianza de los mismos es función de X , de forma que al aumentar la variable endógena se incrementa también la varianza de los residuos
- Las predicciones que se realicen con un modelo en el que exista heteroscedasticidad no serán eficientes, apareciendo unos intervalos excesivamente amplios, ya que la varianza de la predicción incluye la varianza del término aleatorio y la de los estimadores de los parámetros, y estas no son mínimas.

1.8.2.2. Solución a la heteroscedasticidad

La estimación alternativa al uso de MCO en situaciones de heteroscedasticidad es la utilización de MCG y, por tanto, esta es una estrategia analíticamente

correcta para la solución del problema. No obstante, esto implica conocer el verdadero valor de la matriz sigma de varianzas y covarianzas, situación que, en la práctica, no es habitual.

1.8.3. Autocorrelación

La autocorrelación se define, como la existencia de correlación entre perturbaciones aleatorias correspondientes a períodos distintos:

$$Cov(u_i, u_j) = \sigma_{ij} \neq 0 \quad \forall i \neq j$$

El problema de la autocorrelación se denomina también frecuentemente de “correlación serial”. El término de autocorrelación se asocia fundamentalmente con series temporales. Una de las cosas de la aparición de correlación serial reside en la propia naturaleza del término de perturbación, que recoge aquellas variables que se mueven conjuntamente a lo largo del tiempo, pero que de hecho no son individualmente relevantes para la explicación de la variable dependiente. Asimismo, la presencia de autocorrelación se puede explicar por razones generales como el crecimiento y los movimientos cíclicos de la economía. También puede ser motivada por los

errores de especificación del modelo, bien por omisión de alguna variable o por una especificación funcional incorrecta.

1.8.3.1. Detección de autocorrelación

El test más empleado para detectar la autocorrelación en un modelo es el estadístico Durbin- Watson. Este test es una medida de correlación serial entre los residuos. El estadístico Durbin-Watson se denota mediante la letra d y se expresa como sigue:

$$d = \frac{\sum_{t=2}^n \left(\hat{u}_t - \hat{u}_{t-1} \right)^2}{\sum_{t=1}^n \hat{u}_t^2}$$

Para explicar intuitivamente el uso del estadístico d conviene recordar que para muestras grandes, la ecuación presentada equivale a:

$$d = 2 - 2\hat{\rho}$$

Donde $\hat{\rho}$ es el coeficiente de autocorrelación de los residuos, esto es

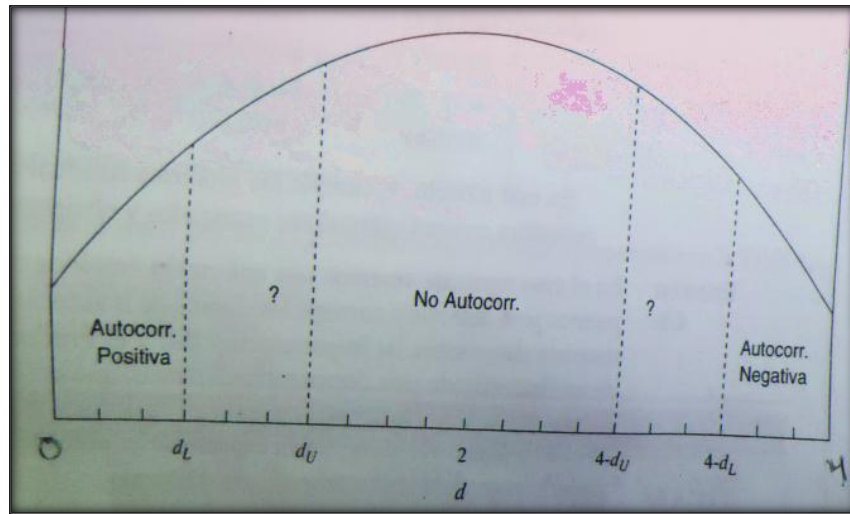
$$\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$$

Así pues caben tres posibilidades:

- Si $\rho = 0$, es decir, si no existe autocorrelación entre los residuos, el estadístico d toma valores entorno a 2
- Si $\rho = 1$, es decir si existe autocorrelación positiva perfecta, el estadístico toma valores cercanos a 0
- Si $\rho = -1$, lo que indica la existencia de autocorrelación negativa perfecta, el estadístico estará próximo a $d=4$.

Por lo tanto, una regla práctica para contrastar, de forma aproximada, si existe autocorrelación entre los residuos, consiste en calcular la muestra con la que estamos trabajando el estimado del estadístico d y comprobar si está cercano a 2. Cuando el estadístico Durbin Watson es significativamente distinto de 2 se concluirá que existe autocorrelación.

En términos gráficos, el estadístico de Durbin Watson (d) se puede representar de la forma siguiente



Donde

d_L = Límite inferior

d_U = Límite superior

1.8.3.2. Solución al problema de la autocorrelación

Para evitar las consecuencias de la autocorrelación caben distintas posibilidades, según el motivo que lo genera. Una fuente muy habitual de generación a este problema se debe a una correcta especificación del modelo. Tratando de evitarlo, el paso previo consiste en incluir en la especificación determinadas variables y/o revisar la forma funcional del modelo.

- Un método para eliminar la autocorrelacion es agregar a la función de regresión una variable omitida que explique la asociación en la respuesta de un periodo al siguiente.
- Otro método involucra la noción general de diferenciar.

1.9. Medición del error de pronóstico

1.9.1. Notación básica de pronósticos

Las técnicas cuantitativas de pronósticos con frecuencia implican datos de series de tiempo, se desarrolló una notación matemática para referirse a cada periodo específico. La letra Y , se utiliza para indicar una variable de serie de tiempo, a menos que haya más de una variable implicada. El periodo asociado con una observación se muestra como un subíndice. Así, Y_t se refiere al valor de una serie de tiempo en el periodo t .

La notación matemática también debe ser desarrollada para distinguir entre un valor real de una serie de tiempo y un valor pronosticado. Un \hat{Y}_t (^ sombrero) se le coloca encima de un valor para indicar que se está pronosticando con

frecuencia se juzga mediante la comparación de la serie original Y_1, Y_2, \dots con la serie de valores pronosticados $\hat{Y}_1, \hat{Y}_2, \dots$

Se ha creado diversos métodos para resumir los errores generados por una técnica de pronósticos específica. La mayoría de estas medidas son el promedio de alguna función de la diferencia entre los valores reales y de pronósticos. Estas diferencias se conocen como residuales.

Residual o error de un pronóstico:

$$u_t = Y_t - \hat{Y}_t$$

Donde

u_t = error de pronóstico en el periodo t

Y_t = valor real en el periodo t

\hat{Y}_t = valor de pronóstico para el periodo t

1.9.2. Medidas de precisión

Error porcentual absoluto medio (MAPE por sus siglas en inglés, mean absolute percentage error): se calcula al encontrar el error absoluto en

cada periodo, dividiendo entre el valor real observado para ese periodo y luego promediando los errores porcentuales absolutos. Este método es útil cuando el tamaño o magnitud de la variable del pronóstico es importante para evaluar la precisión del pronóstico.

Especialmente útil cuando los valores de Y_t son grandes. También se puede usar para comparar la precisión de las mismas o diferentes técnicas en dos series totalmente distintas.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t}$$

Desviación absoluta media (MAD por sus siglas en inglés, mean absolute deviation): mide la precisión del pronóstico al promediar las magnitudes de los errores de pronóstico. es útil cuando se quiere medir el error de pronóstico en las mismas unidades que la serie original.

$$MAD = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|$$

Error cuadrático medio (MSE por sus siglas en inglés, mean squared error): es otro método para evaluar una técnica de pronóstico. Cada error

de pronóstico se eleva al cuadrado; luego, se suman y se dividen entre el número de observaciones.

Este método penaliza los errores grandes de pronóstico debido a que los errores se elevan al cuadrado, lo cual es importante; una técnica que produce errores moderados podría ser preferible a una que, por lo general, tiene errores pequeños, pero que en ocasiones produce errores muy grandes.

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2$$

Error porcentual medio (MPE por sus siglas en inglés, mean porcentaje error): determina si un método de pronóstico tiene sesgo. Si el método de pronóstico no tiene sesgo, el MPE producirá un número cercano a cero. Si el resultado es un porcentaje negativo, el método sobreestima de forma consistente, y si el resultado es un porcentaje alto positivo, el método subestima consistentemente.

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)}{Y_t}$$

Las medidas de precisión de pronóstico mencionadas, se utilizan para comparar la precisión de dos (o más) técnicas diferentes, medir la utilidad o confiabilidad de una técnica específica y ayudar a buscar una técnica óptima.

1.10. Modelo autorregresivo (AR)

Un modelo autorregresivo de cualquier orden p adopta la siguiente forma:

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \varepsilon_t$$

Donde:

Y_t = Variable endógena en el tiempo t

$Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ = Variables endógenas en los retrasos $t-1, t-2, \dots, t-p$, respectivamente, estas Y desempeñan la función de variables exógenas.

$\Phi_0, \Phi_1, \Phi_2, \dots, \Phi_p$ = Coeficientes que serán estimados

ε_t = Término de error en el tiempo t que representa los efectos de la variable que no explica el modelo, los supuestos acerca del término del error son las mismas que las del modelo de regresión estándar.

Los modelos autorregresivos son apropiados para datos de series de tiempo estacionarios y que tienen un coeficiente Φ_0 que se relaciona con el nivel constante de la serie. Si los datos varían alrededor de la serie o se expresan como desviaciones de la media $Y_t - \bar{Y}$, no se requiere el coeficiente Φ_0 .

1.11. Modelo de media móvil (MA)

Un modelo de media móvil de orden q adopta la forma:

$$Y_t = \mu + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \dots - \omega_q \varepsilon_{t-q}$$

Donde:

Y_t = variable endógena (dependiente en un tiempo t)

μ = valor promedio que permanece constante en el proceso.

$\omega_1, \omega_2, \dots, \omega_q$ = coeficientes que serán estimados

ε_t = término de error que representa los efectos de las variables exógenas por el modelo; los supuestos acerca del término de error son las mismas que las del modelo de regresión estándar.

$\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ = errores en periodos anteriores al tiempo t, incorporados en la endógena Y_t

Esta ecuación es similar a la ecuación de modelos autorregresivos, tomando en cuenta que la variable endógena Y cambia por ε . Los modelos de media móvil (MA) proporcionan pronósticos de Y , con base en una combinación lineal de un número finito de errores pasados, mientras que los modelos autorregresivos (AR) pronostican a Y como una función lineal de un número finito de valores anteriores de Y .

1.12. Modelos de medias móviles autorregresivos (ARMA)

Para estos modelos es conveniente usar la notación ARMA (p,q) donde p equivale al orden de la parte autorregresiva y q es el orden de la parte de media móvil. Un modelo ARMA (p,q) tiene la forma general:

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \dots - \omega_q \varepsilon_{t-q}$$

Los modelos ARMA (p,q) pueden describir una amplia variedad de comportamientos para las series de tiempo estacionarias, los pronósticos generados por un modelo ARMA(p,q) dependerán de los valores actuales y anteriores de la respuesta Y, así como de los valores pasados y actuales de los errores.

Los números de los términos autorregresivos y de media móvil (órdenes p y q) en un modelo ARMA se determinan a partir de los patrones de las autocorrelaciones de la muestra y de las autocorrelaciones parciales. En la práctica rara vez p y q exceden a 2.

1.13. Modelos ARIMA

Los modelos de media móvil autorregresivo integrado (ARIMA, por sus siglas en inglés). Cada una de las tres partes de las siglas ARIMA se le denomina componente y modela un comportamiento distinto de la serie.

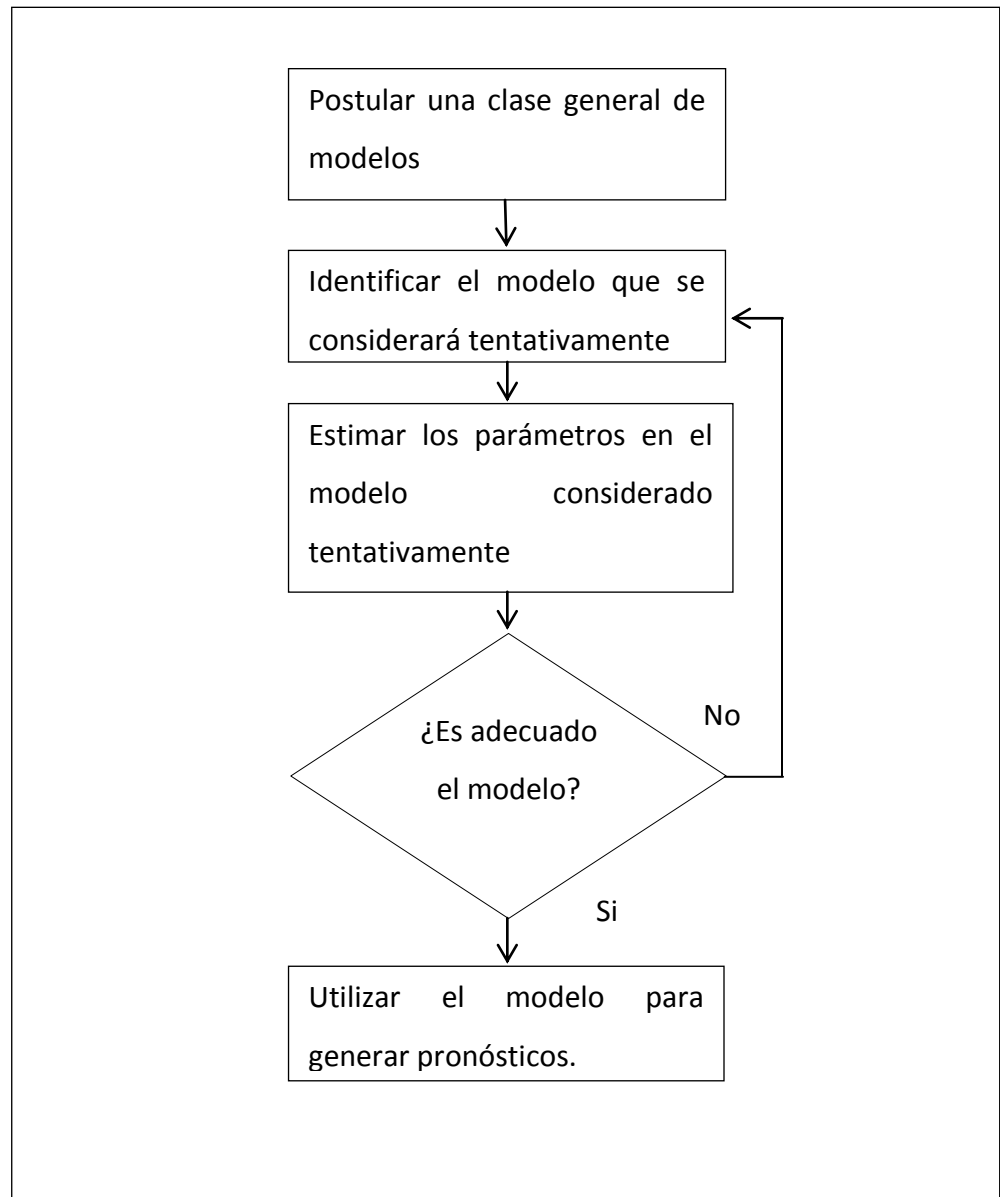
Los modelos ARIMA son una clase de modelos lineales que tienen la capacidad de operar sobre series de tiempo estacionarias o no estacionarios. Donde los procesos estacionarios varían en torno a un nivel fijo en tanto que los procesos no estacionarios no tienen un nivel promedio constante natural.

Estos modelos no involucran a las variables independientes en su construcción. En, cambio emplea información que se encuentra en la serie misma para generar los pronósticos. Dependen mucho de los patrones de autocorrelación que existen entre los datos.

La selección inicial de un modelo ARIMA se basa en el examen de una gráfica de la serie de tiempo y un examen de su autocorrelación para diversos retrasos. De manera particular, el patrón de la autocorrelación de la muestra, calculado a partir de la serie de tiempo, coincide con el patrón de autocorrelación ya conocido que se asocia a un modelo ARIMA específico. Este acoplamiento se asocia para las autocorrelaciones y autocorrelaciones parciales.

1.13.1. Aplicación de una estrategia para la construcción de un modelo (ARIMA)

Figura 1: Estrategia para la construcción de un modelo ARIMA



Fuente: Pronósticos en los negocios de Hanke

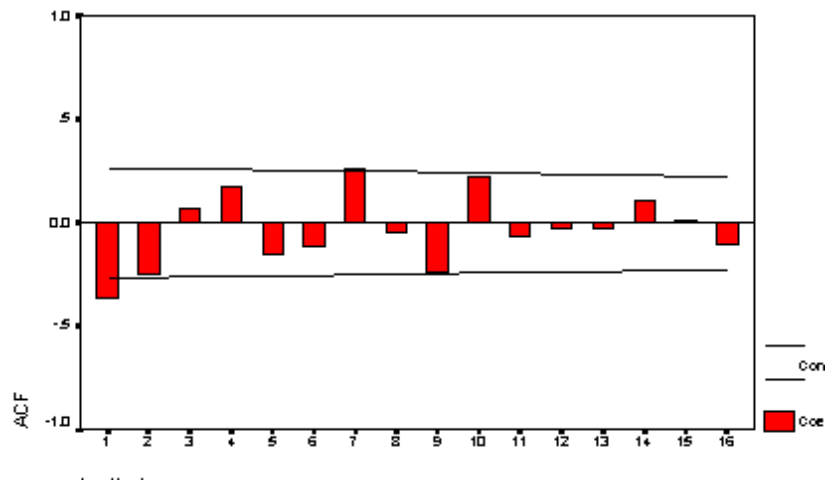
Elaborado por: Laura Rochina

El método Box-Jenkins utiliza una estrategia iterativa para la construcción de modelos que consiste en seleccionar un modelo inicial mediante la estimación de los coeficientes del modelo y el análisis de los residuales

Paso 1: Identificación del modelo.

El primer paso en la identificación del modelo es determinar si la serie es estacionaria o no; como se observa el siguiente correlograma.

Gráfica 6: Correlograma de una serie de tiempo



Si la serie no es estacionaria Y_t , con frecuencia puede convertirse en una serie estacionaria ΔY_t , al tomar sus diferencias. Es decir, la serie se reemplaza por una serie de diferencias $\Delta Y_t = Y_t - Y_{t-1}$. Entonces, se especifica un modelo ARMA para la serie de las diferencias.

$$\Delta Y_t = \phi_1 \Delta Y_{t-1} + \varepsilon_t - \omega_1 \varepsilon_{t-1}$$
$$Y_t - Y_{t-1} = \phi_1 (Y_t - Y_{t-1}) + \varepsilon_t - \omega_1 \varepsilon_{t-1}$$

En efecto, el analista está modelando los cambios ΔY_t en lugar de los niveles.

En algunos casos podría ser necesario tomar las diferencias de las diferencias antes de obtener los datos estacionarios. Cuando se toma una diferencia simple dos veces los datos estacionarios son:

$$\Delta^2 Y_t = \Delta(\Delta Y_t) = \Delta(Y_t - Y_{t-1}) = Y_t - Y_{t-1} - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$$

Se toman diferencias hasta que la gráfica de los datos indica que la serie varían alrededor de un nivel fijo y las autocorrelaciones de la muestra desaparecen con rapidez. El número de diferencias requerido para lograr un estado estacionario se denota por d.

A los modelos para las series que no son estacionarias se les llama modelos de medias móviles integrados autorregresivos y se denota como ARIMA (p,d,q). En este caso p, indica el orden de la parte autorregresiva, d indica el orden de la diferencia y q, el orden de la parte de media móvil.

Si la serie original es estacionaria entonces $d=0$ y los modelos ARIMA se reducen a modelos ARMA.

Una vez que se ha obtenido una serie estacionaria, el analista debe identificar la forma del modelo que habrá de utilizar.

1. Por medio de las gráficas de autocorrelación parcial y autocorrelacion definimos los parámetros p, q respectivamente.
2. Cada modelo ARIMA cuenta con una serie única de autocorrelaciones y autocorrelaciones parciales, debemos tener la capacidad de identificar los valores que se obtienen a partir de la muestra.

Es probable que haya una cierta ambigüedad al determinar un modelo ARIMA apropiado a partir de los patrones que provienen de las autocorrelaciones y autocorrelaciones parciales de la muestra. De esta manera, la selección del modelo inicial deberá ser considerando como tentativa.

Paso 2: Estimación de los modelos.

Los parámetros en los modelos ARIMA se estiman al minimizar la suma de los cuadrados en los errores de ajuste. En general, estos estimados deben obtenerse mediante un procedimiento lineal de mínimos cuadrados.

Además se calcula el error cuadrado medio de los residuales, un estimado de los residuales de la varianza ε_t

Paso 3: Evaluación del modelo

Básicamente un modelo es adecuado si sus residuales no pueden ser utilizados para mejorar los pronósticos. Es decir, los residuales deben ser aleatorios.

- Muchas gráficas de los residuales que son útiles para el análisis de regresión pueden desarrollarse para los residuales de un modelo ARIMA. De manera particular, son útiles un histograma, gráficos de probabilidad normal y una gráfica de secuencias de tiempo.
- Las autocorrelaciones residuales individuales, deberán ser pequeñas y por lo general estar dentro de $\pm 2/\sqrt{n}$; donde n es el número de observaciones. Las autocorrelaciones significativas en retrasos cortos o estacionales sugieren que el modelo no es bueno y que se debe elegir un modelo nuevo o modificado.
- Como un grupo, las autocorrelaciones residuales deberán ser coherentes con aquellas producidas por los errores aleatorios.

Una prueba chi cuadrada (χ^2) que se basa en la estadística Ljung-Box Q proporciona una revisión global de la pertinencia del modelo. Esta prueba considera las dimensiones de las autorrelaciones residuales como un grupo. La estadística de prueba Q es:

$$Q_m = n(n+2) \sum_{k=1}^m \frac{r_k^2(\ell)}{n-k}$$

Donde:

$r_k(\ell)$ = autocorrelación residual en el retraso k

n = número de residuales

k = retraso de tiempo

m = número de retrasos de tiempo que habrán de ser evaluados

si el valor p asociado con la estadística Q es pequeño (por ejemplo, un valor $p < 0,05$), se considera que el modelo es inadecuado

Paso 4: Realización de pronósticos con el modelo.

Después de que se ha encontrado un modelo adecuado, se puede llevar a cabo los pronósticos para un periodo, o varios, en el futuro. También pueden construirse intervalos de predicción con base en los pronósticos.

En general para un nivel de confianza determinado, mientras más largo sea el tiempo, guía del pronóstico mayor será el intervalo de predicción. Esto es sensato porque se espera que la incertidumbre para el pronóstico de un valor lejano. Es tedioso calcular los pronósticos y los intervalos de predicción; por eso es mejor hacerlo mediante la computadora.

A medida que se tiene más datos disponibles, se puede usar el mismo modelo ARIMA para generar pronósticos revisados que procedan de otro origen de tiempo.

Si el patrón de la serie parece cambiar con el tiempo, los nuevos datos podrían usarse para volver a estimar los parámetros del modelo, de ser necesario, desarrollar un modelo completamente nuevo.

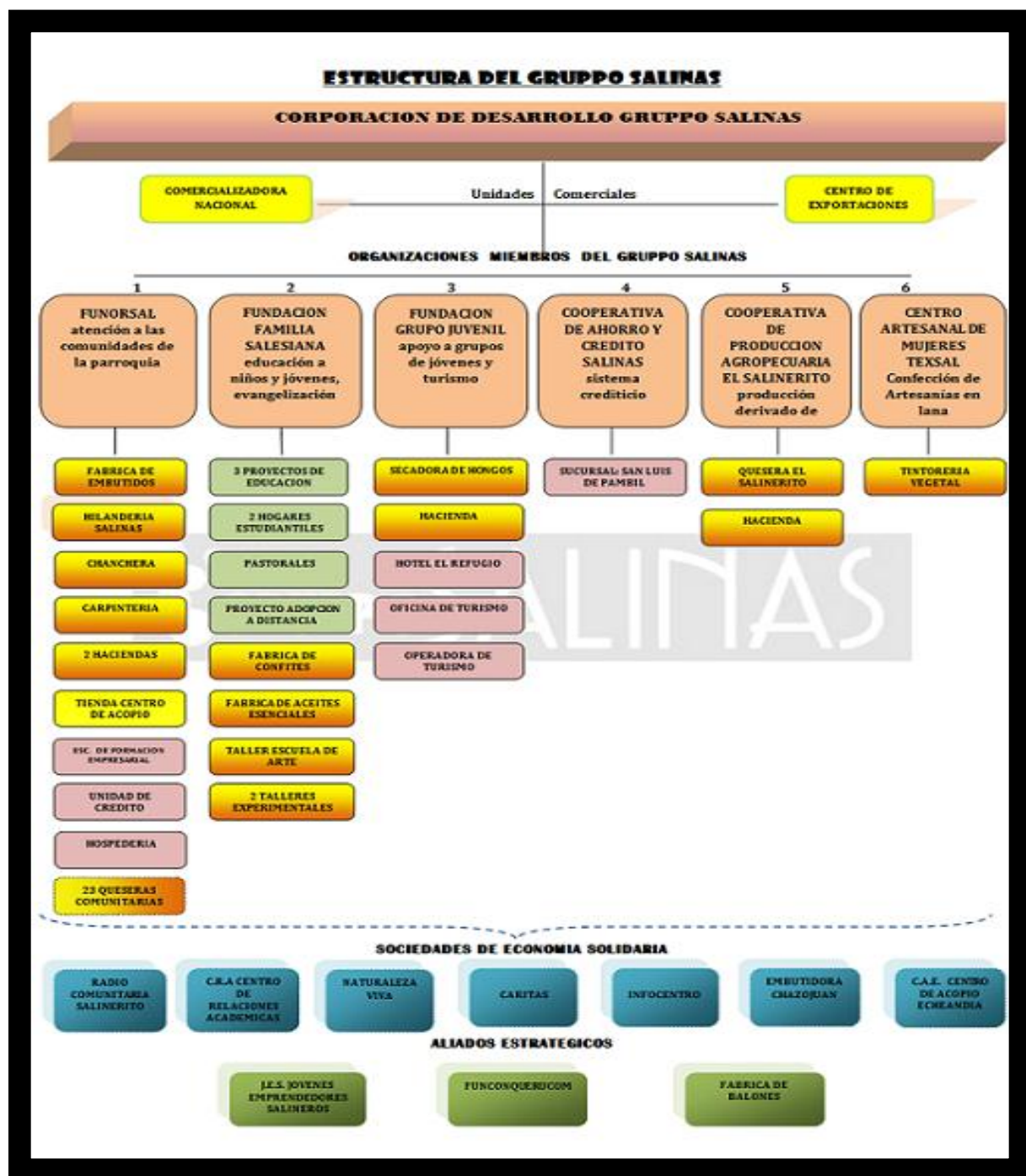
1.14. Gruppo salinas

El GRUPPO SALINAS (jurídicamente aprobada el 26 de noviembre del 2006) constituye una instancia corporativa que representa los intereses comunes y específicos de la población Salinera, y de las distintas instituciones miembros; es la llamada a articular procesos y apoyar iniciativas locales para fortalecer el proceso de trabajo comunitario. Es un Organismo Técnico, capaz de tomar el liderazgo parroquial y regional conjuntamente con los gobiernos locales en los

aspectos de equidad de género, participación ciudadana, protección del medioambiente, equilibrio generacional, productivo y de comercialización. Todo esto se logra con la intervención de actores públicos y privados, capacitados y especializados que comparten sus conocimientos con el equipo local y las comunidades, organizaciones participantes, generando a corto y mediano plazo capacidades locales, para continuar con la tarea emprendida.

El Gruppo Salinas, como ente corporativo, está integrado por seis miembros que a su vez son organizaciones salineras dedicadas a actividades sociales y productivas.

Figura 2: Estructura del GRUPPO Salinas



Elaborado por: María Vargas

Fuente: Gruppo Salinas

1.14.1. Hilandería Intercomunal Salinas (H.I.S.)

La hilandería es una empresa de tipo industrial dotada de máquinas importadas de Canadá. Se encarga de la transformación de la lana de oveja, llama y alpaca en un hilo de calidad que se vende en el mercado nacional (además de utilizarse localmente por la empresa de tejidos). El abastecimiento de lana se hace a través de un sistema societal¹ integrado por diferentes comunidades campesinas ubicadas por todo el país, que permite eliminar los intermediarios y favorece a otras comunidades campesinas del Ecuador.

Varias personas trabajan en esta fábrica, distribuidas en horarios convenientes para el obrero, que les permiten trabajar cuatro días y descansar cuatro. En este tiempo se ocupan de la ganadería y la agricultura, con lo que aumentan sus ingresos.

¹ “Sistema Societal”, sistema de sociedad donde se busca el beneficio de todos sus integrantes

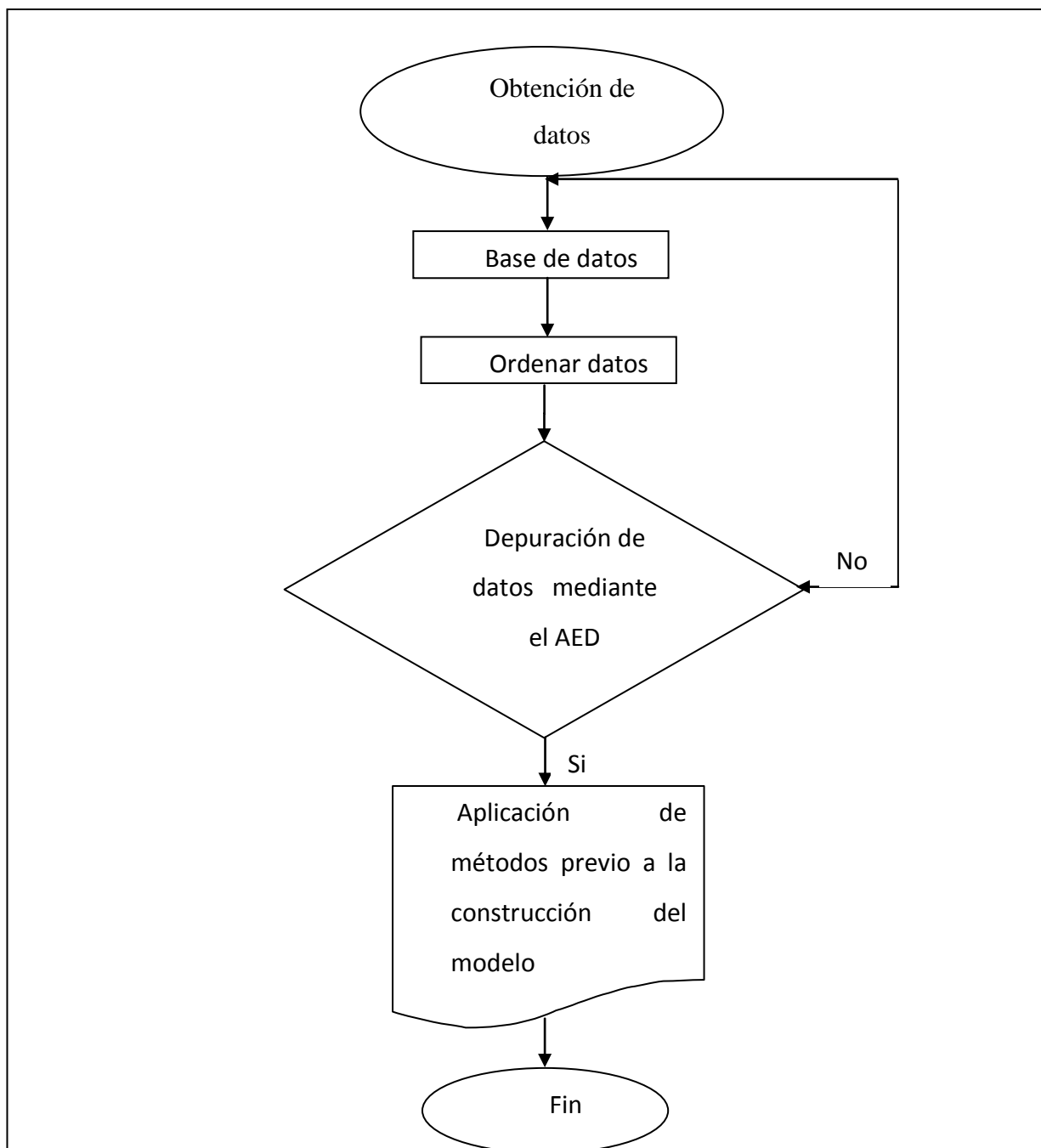
CAPÍTULO

II

2. METODOLOGÍA

2.1. Recolección de datos

Figura 3: Recolección de datos



Elaborado por: Laura Rochina

Fuentes de información.- Son todos aquellos medios de los cuales procede la información para la elaboración del presente proyecto de tesis, que satisfacen las necesidades de conocimiento de la producción de hilo, que posteriormente será utilizado para lograr los objetivos esperados. De acuerdo a su origen se clasifican en:

- Fuentes primarias.
- Fuentes secundarias.

Fuentes primarias.- Son aquellas en las que los datos provienen directamente de la población o muestra, en este los registros de la producción de hilo de la empresa, mientras que las fuentes secundarias son aquellas que parten de datos pre-elaborados, como la información que se obtiene de FUNORSAL.

Fuentes secundarias.- Para ser utilizadas son analizadas bajo 4 preguntas básicas:

- ¿Es pertinente? cuando la información se adapta a los objetivos.
- ¿Es obsoleta? cuando ha perdido actualidad, por eso en esta investigación se toma datos mensuales desde el año 2008.

- ¿Es Fidedigna? cuando la veracidad de la fuente de origen no es cuestionada, en este caso la información se obtendrá de los registros mensuales de la producción de hilo de la empresa.
- ¿Es digna de Confianza? si la información ha sido obtenida con la metodología adecuada y honestidad necesaria, con objetividad, naturaleza continuada y exactitud

Observación directa.- Es cuando se toma directamente los datos sin necesidad de realizar algún medio para obtener información

2.1.1. Análisis exploratorio de datos (AED)

El Análisis Exploratorio de Datos (AED) es un conjunto de técnicas estadísticas cuya finalidad es conseguir un entendimiento básico de los datos y de las relaciones existentes entre las variables analizadas. Para conseguir este objetivo el A.E.D. proporciona métodos sistemáticos sencillos para organizar y preparar los datos, detectar fallos en el diseño y recogida de los mismos, tratamiento y evaluación de datos ausentes (missing), identificación de casos atípicos (outliers) y comprobación de los supuestos subyacentes en la mayor parte de las técnicas multivariantes (normalidad, linealidad, homocedasticidad).

2.1.1.1. Etapas del AED

Para realizar un AED conviene seguir las siguientes etapas:

- 1) Preparar los datos para hacerlos accesibles a cualquier técnica estadística.
- 2) Realizar un examen gráfico de la naturaleza de las variables individuales a analizar y un análisis descriptivo numérico que permita cuantificar algunos aspectos gráficos de los datos.
- 3) Realizar un examen gráfico de las relaciones entre las variables analizadas y un análisis descriptivo numérico que cuantifique el grado de interrelación existente entre ellas.
- 4) Evaluar, los supuestos básicos subyacentes a muchas técnicas estadísticas como, por ejemplo, la normalidad, linealidad y homocedasticidad.
- 5) Identificar los posibles casos atípicos (outliers).

2.1.1.2. Preparación de los datos

El primer paso en un AED es hacer accesible los datos a cualquier técnica estadística. Ello conlleva la selección del método de entrada (por teclado o importados de un archivo) y codificación de los datos así como la de un paquete estadístico adecuado para procesarlos.

La inmensa mayoría de los paquetes estadísticos permite realizar manipulaciones de los datos previos a un análisis de los mismos. Algunas operaciones útiles son las siguientes:

- Combinar conjuntos de datos de dos archivos distintos.
- Seleccionar subconjuntos de los datos
- Dividir el archivo de los datos en varias partes
- Transformar variables
- Ordenar casos
- Agregar nuevos datos y/o variables

- Eliminar datos y/o variables
- Guardar datos y/o resultados

2.1.1.3. Análisis estadístico unidimensional

Una vez preparados los datos, el segundo paso de un AED consiste en realizar un análisis estadístico gráfico y numérico de las variables del problema con el fin de tener una idea inicial de la información contenida en el conjunto de datos, así como detectar la existencia de posibles errores en la codificación de los mismos. El tipo de análisis a realizar depende de la escala de medida de la variable analizada

2.1.2. Estadísticos para el análisis exploratorio de datos (AED)

- **Media:** Nos dará un resumen común al promediar los datos

$$\bar{X} = \frac{\sum X}{n}$$

Donde

\bar{X} = media de la muestra

$\sum X$ = suma de los n valores de la muestra

n = tamaño de la muestra

- **Varianza muestral:** Se define como el cuadrado de las desviaciones estándar de un conjunto de medidas

$$S^2 = \frac{\sum (X - \bar{X})^2}{n-1}$$

- **Desviación estándar:** Ayudará a medir la dispersión de los n datos con respecto a cada una de las variables en estudio

$$S = \sqrt{S^2}$$

- **Coeficiente de asimetría:** Mide la asimetría de los datos respecto a su centro

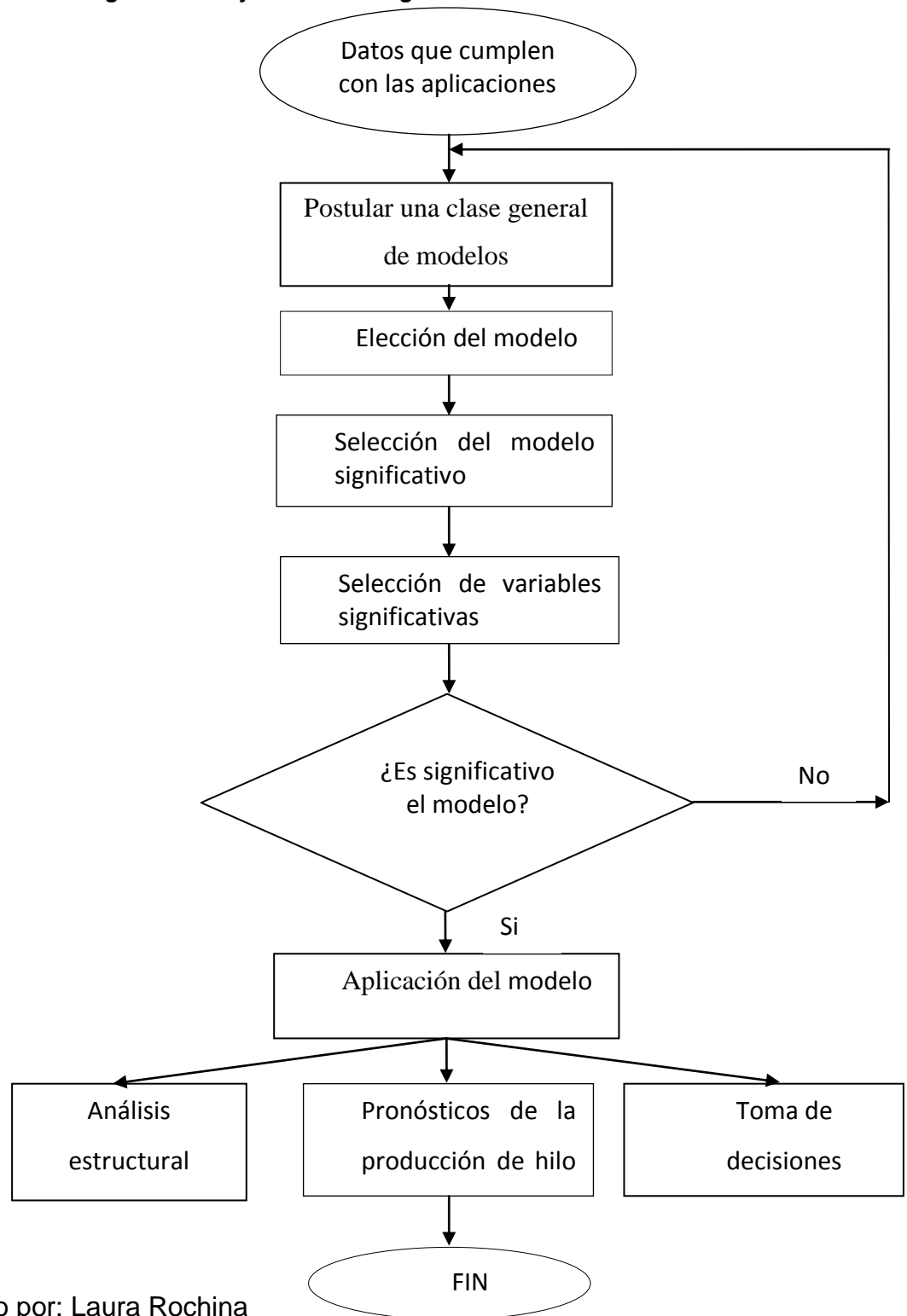
$$A = \frac{1}{n} \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{S^3}$$

- **Coeficientes de Kurtosis:** Es una medida que dice qué tan puntiaguda es la distribución de probabilidad.

$$CK = \frac{1}{n} \frac{\sum (x - \bar{x})^4}{S^4}$$

2.2. MODELACION DE DATOS

Figura 4: Diagrama de flujo de la estrategia de la construcción del modelo



Elaborado por: Laura Rochina

2.2.1. Modelos de pronósticos

Luego de haber formado la base de datos y aplicado las técnicas de depuración se puede escoger ya un modelo general con todas las variables existentes e influyentes a la producción de hilo aplicando las técnicas de exploración de patrones de datos de series de tiempo, métodos de modelación y pronósticos:

2.2.2. Exploración de patrones de datos de series de tiempo

Una serie de tiempo consta de datos que reúnen, registran u observan sobre incrementos sucesivos de tiempo. Se requiere un enfoque sistemático para analizarlas. A continuación se presenta en la tabla1 una breve descripción sobre los patrones de series de tiempo.

Tabla 1: Componentes de un Serie de tiempo

Componente	Descripción
Tendencia	Es el componente de largo plazo que representa el crecimiento o disminución en la serie sobre un periodo amplio.
Cíclico	Es la fluctuación en forma de onda alrededor de la tendencia.
Estacional	Es un patrón de cambio que se repite a sí mismo año tras año.
Aleatorio	Mide la variabilidad de las series de tiempo después de retirar los otros componentes.

Elaborado por: Laura Rochina

2.2.3. Selección de variables significativas

Aunque la especificación de un modelo es una labor eminentemente teórica, las reglas prácticas se justifican porque facilitan la selección de variables. Se tiene en cuenta que, si bien la inclusión de una variable determinada debe siempre tener un claro soporte técnico. Los procedimientos para seleccionar las variables dependientes que deben entrar en el modelo son los siguientes:

- Una vez que disponemos de variables, algunas de las cuales pueden estar medidas de diversas formas, se calculará la matriz de correlación para obtener una visión global de las interrelaciones existentes entre las mismas. La matriz de correlación nos indicará las variables que están más correlacionadas con la variable dependiente, además la posible existencia de multicolinealidad entre las variables explicativas.
- Las variables elegidas deben ser coherentes con la teoría económica que oriente la formulación del modelo.
- Los signos de los coeficientes deben responder al conocimiento teórico a priori.
- El estadístico F y los estadísticos t deben rechazar las hipótesis nulas que

cada test contrasta.

- El comportamiento de los residuos debe ser aleatorio para evaluar el modelo.

2.2.4. Selección de un modelo significativo

2.2.4.1. Prueba F

La prueba F para la significancia del modelo planteado, las hipótesis:

1. Establecimiento de hipótesis.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{al menos una } \beta_j \neq 0$$

2. Selección de un nivel de significancia

Generalmente se toman valores del orden 0.01, 0.05 ó 0.10 ó entre 0.01 y 0.10

3. Descripción del tamaño de la muestra y cálculos de estadístico(s)

Se toma en cuenta el tamaño de la muestra y aspectos de las distribuciones muestrales

Son probadas por la razón F

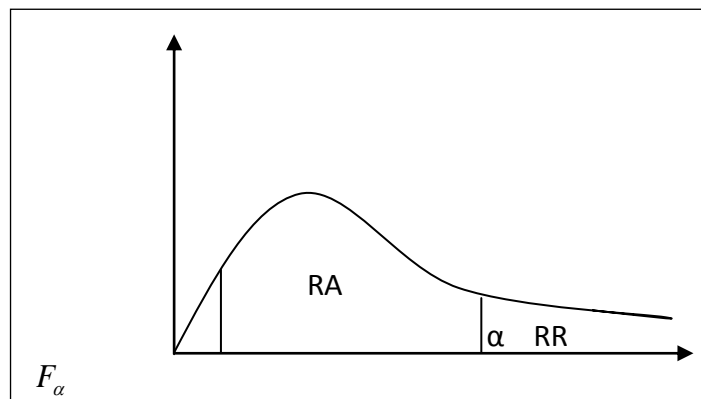
$$F = \frac{MSR}{MSE}$$

Con $df = k, n - k - 1$ grados de libertad.

$$F > F_{\alpha}$$

4. Especificación de las regiones de aceptación y rechazo

En un nivel de significancia de α la región de rechazo es:



5. Decisión estadística

Para tomar una decisión, se compara el valor de la estadística calculada con el valor crítico, o se observa si este valor, cae en la región de aceptación o rechazo y luego dependiendo de esta comparación se acepta o se rechaza la hipótesis nula H_0 .

2.2.4.2. Medidas de bondad de ajuste

Coeficiente de determinación (R^2): Representa la razón de la variación en la variable endógena por su relación con las exógenas.

$$R^2 = \frac{S_{\hat{Y}}^2}{S_Y^2}$$

Error estándar: Viene expresada en la unidad de medida de la variable endógena

$$S_{\hat{u}}^2 = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2} = \frac{\sum \hat{u}_i^2}{n-2}$$

Error o residuo: El análisis de los residuos puede ser gráficamente o analíticamente.

$$e = Y - \hat{Y}$$

CAPÍTULO

III

3. Elaboración del modelo econométrico para el pronóstico de la producción de hilo

3.1. Matriz de datos

Para el desarrollo de la investigación se recolectó información de la producción (variable endógena) y variables exógenas, de los archivos mensuales desde el año 2008 hasta el mes de mayo del 2012 de la empresa.

Tabla 2: Matriz de datos.

	Meses	MESES COD	Producción	materia prima	materia prima perdida	costo de insumos	costo de materia prima	Mano de obra	piola plastica
2008	Enero	1	4.492	11.000	985	978	2.109	780	19,05
	Febrero	2	5.980	13.432	1.284	325	3.905	824	18,95
	Marzo	3	6.352	14.058	3.601	687	6.352	832	20,54
	Abril	4	5.045	14.879	4.198	324	3.791	890	22,95
	Mayo	5	6.530	15.098	3.372	780	3.120	830	19,30
	Junio	6	7.076	17.406	5.023	1.954	3.125	780	16,90
	Julio	7	5.253	14.631	3.521	1.098	2.941	801	19,06
	Agosto	8	9.753	18.501	5.934	897	3.682	890	20,32
	Septiembre	9	2.241	10.794	1.974	453	2.574	905	17,04
	Octubre	10	5.023	16.503	3.420	231	3.015	970	17,03
	Noviembre	11	6.091	13.863	4.072	984	2.084	867	19,04
	Diciembre	12	5.119	12.120	1.153	652	3.965	960	18,56
2009	Enero	13	7.457	12.345	1.798	1.110	3.901	1.678	39,87
	Febrero	14	5.605	12.143	4.036	1.317	4.880	1.043	28,09
	Marzo	15	8.023	14.641	3.087	2.102	4.330	1.234	30,98
	Abril	16	6.892	12.567	5.721	970	4.081	1.104	25,05
	Mayo	17	8.005	16.328	2.659	845	4.033	1.078	19,00
	Junio	18	5.143	15.987	632	325	3.701	1.576	45,09
	Julio	19	6.687	13.648	2.756	1.203	4.965	1.365	28,04
	Agosto	20	5.303	14.732	3.641	967	3.911	1.278	23,43
	Septiembre	21	6.501	11.658	1.211	324	4.120	1.304	18,90
	Octubre	22	4.160	12.310	2.740	1.086	4.125	1.284	19,32
	Noviembre	23	5.318	11.089	731	985	3.912	1.309	20,21
	Diciembre	24	3.246	12.700	1.802	1.320	4.071	1.190	21,04
2010	Enero	25	6.524	15.270	2.902	1.212	4.260	2.055	114,45
	Febrero	26	5.404	13.426	3.571	1.986	4.792	1.676	60,56
	Marzo	27	5.070	15.599	5.696	1.239	3.702	1.370	50,43
	Abril	28	5.967	11.539	7.266	1.108	4.320	2.002	20,89
	Mayo	29	5.241	14.074	3.923	786	4.475	2.069	76,30
	Junio	30	9.101	19.488	666	675	5.901	1.973	50,23
	Julio	31	3.675	9.129	1.636	505	5.840	2.018	20,61
	Agosto	32	5.465	10.833	3.248	1.301	7.982	1.367	18,08
	Septiembre	33	5.392	9.105	2.450	1.114	3.873	2.128	19,09
	Octubre	34	4.955	7.517	2.644	2.105	7.745	1.978	18,43
	Noviembre	35	4.272	8.502	958	355	5.361	2.069	18,02
	Diciembre	36	6.098	12.540	2.098	1.034	3.965	1.967	40,23
2011	Enero	37	8.085	11.181	1.628	1.313	4.129	2.358	121,40
	Febrero	38	7.866	14.002	5.253	2.268	4.821	2.396	87,97
	Marzo	39	9.580	14.672	4.648	1.185	19.195	2.570	15,38
	Abril	40	6.840	15.171	4.088	853	4.316	2.402	90,01
	Mayo	41	6.111	19.100	5.019	1.005	5.456	2.582	78,69
	Junio	42	8.008	13.020	647	1.740	3.114	2.582	24,34
	Julio	43	6.942	19.556	4.897	709	4.106	2.432	69,63
	Agosto	44	7.130	19.532	4.437	267	5.484	2.432	38,00
	Septiembre	45	7.170	12.179	3.018	883	3.924	2.432	65,68
	Octubre	46	6.793	17.260	4.132	950	4.089	2.342	60,89
	Noviembre	47	6.097	15.388	5.165	684	5.957	2.432	65,20
	Diciembre	48	8.353	14.027	944	208	4.894	2.432	59,58
2012	Enero	49	9.556	12.243	1.047	3.101	5.940	2.490	44,56
	Febrero	50	7.389	13.896	2.054	2.075	4.065	2.387	59,75
	Marzo	51	8.156	14.532	3.209	1.986	4.896	2.487	39,75

Fuente: Archivos de la Empresa Intercomunal Salinas

Elaborado por: Laura Rochina

Donde tenemos variables endógena y exógenas:

Variable endógena:

Y: Producción de hilo en libras.

Variables exógenas:

X_1 : Materia prima en libras

X_2 : Materia prima perdida en libras

X_3 : Costo de insumos en dólares

X_4 : Costo de materia prima en dólares

X_5 : Mano de obra en dólares

X_6 : Piola plástica en libras

3.2. Análisis exploratorio de datos (AED)

Previo a la construcción del modelo primero se realizó AED para el estudio descriptivo de los datos, y se analizó cada una de las variables, obteniendo estadísticos como la media, desviación estándar, coeficiente de asimetría y coeficiente de variación.

Tabla 3: Estadísticos descriptivos

Estadísticos Variables	Coef. Simetría	Media	Mediana	Varianza	Des. estándar	Coef. Variación
Producción	0,001	6.124	6.121	212	14,56	0,24%
Materia prima	0,166	13.828	13.896	7.718	87,852	1%
Materia prima perdida	0,2989	3.069	3.087	2.659	51,565	2%
Costo de insumos	1,027	1.058	3984	3.801	61,652	6%
Costo de materia prima	4,31	4.770	4.120	6.092	78,051	2%
Mano de obra	-0,118	1.700	1.970	4.183	64,676	4%
Pirola plástica	1,413	39	24	7.092	84,214	216%

Elaborado por: Laura Rochina

- Como podemos observar en la tabla 3, en cuanto a la variable endógena producción con respecto al coeficiente de simetría con un valor de 0,001 lo que nos indica que es aproximadamente simétrica; así también presenta una producción media mensual de 6.124 libras que se aproximan con la mediana de 6.121 libras, se ratifica lo anteriormente mencionado. Al ser el CV $< 10\%$ (su valor es de 0,24%) se puede decir que la producción de hilo es homogénea.
- En el caso de la variable exógena materia prima, presenta un coeficiente de simetría de 0,16 lo que nos indica que tiene una asimetría positiva, también presenta una media mensual de 13.828 libras que no coincide con la mediana de 13.896 libras, entonces decimos que la variable está poco sesgada. Además presenta un CV del 1% lo que nos indica homogeneidad de los datos
- En cuanto a la variable exógena materia prima perdida, con un coeficiente de simetría de 0,298 lo que indica una distribución asimétrica positiva y presenta una media de 3.069 libras mensuales es diferente de la mediana que presenta 3.087 libras, que indica un sesgo positivo y con un CV del 2% nos indica que esta variable es homogénea.

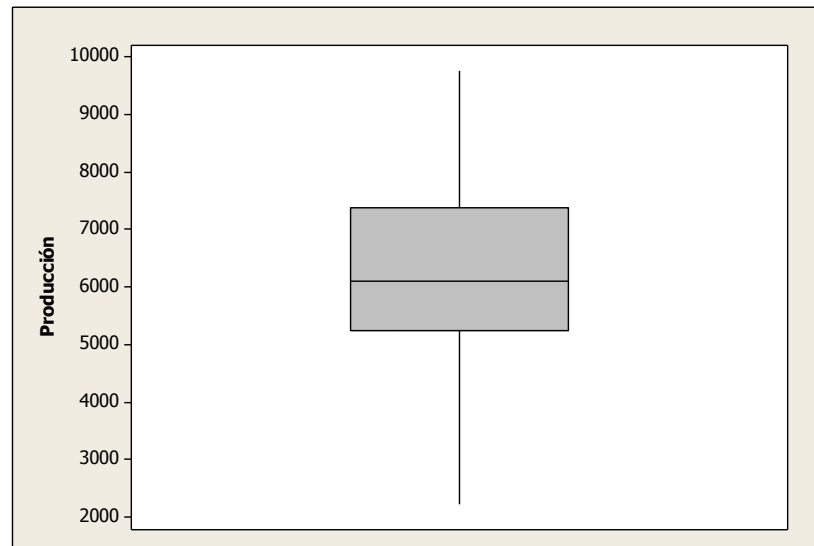
- La variable exógena costo de insumos, presenta asimetría positiva con un coeficiente de 1,027 y un costo medio mensual de 1.058 dólares, el mismo presenta una gran diferencia con la mediana que es un valor de 984 dólares, entonces se ratifica lo anteriormente mencionado. Y el valor del CV de 6% nos indica claramente homogeneidad en la variable
- El costo de materia prima variable exógena, presenta también asimetría positiva con un coeficiente de 4,31 y un costo medio mensual de 4.770 dólares, este se diferencia de la mediana ya que presenta un valor de 4.120 dólares, lo que nos indica un sesgo positivo. Al ser el CV 2% decimos que esta variable exógena presenta homogeneidad.
- En cuanto a la variable mano de obra con respecto a su coeficiente de simetría con un valor de -0,118 decimos que presenta una asimetría negativa, y una media mensual de 1.700 dólares que es diferente a la mediana ya que este presenta un valor de 1.980 dólares, lo que nos indica un sesgo negativo de los datos. Además con un CV del 4% decimos que esta variable es homogénea.

- Piola plástica definida como variable exógena presenta una asimetría positiva con un coeficiente de 1,413, y una media mensual de 39 libras el mismo se diferencia de la mediana que presenta un valor de 24 libras, entonces esta variable tiene un sesgo positivo. Al ser el CV $>100\%$ (su valor es de 240%) se puede decir que la variable piola plástica no es homogénea.
- Al observar el varianza de cada una de las variables encontramos una dispersión muy grande lo cual nos indica que la media y la mediana de cada una de las variables en estudio no son buenos indicadores de los datos.
- Tomando en cuenta el coeficiente de asimetría observamos que todas las variables presentan un sesgo positivo y para comparar entre ellas utilizamos el coeficiente de variación siendo este menor al 10% por lo que nos indica homogeneidad de los datos.

3.2.1. Análisis de datos atípicos

Previo a la aplicación de técnicas estadísticas para la producción de hilo se realizó el estudio de datos atípicos, por medio del análisis gráfico:

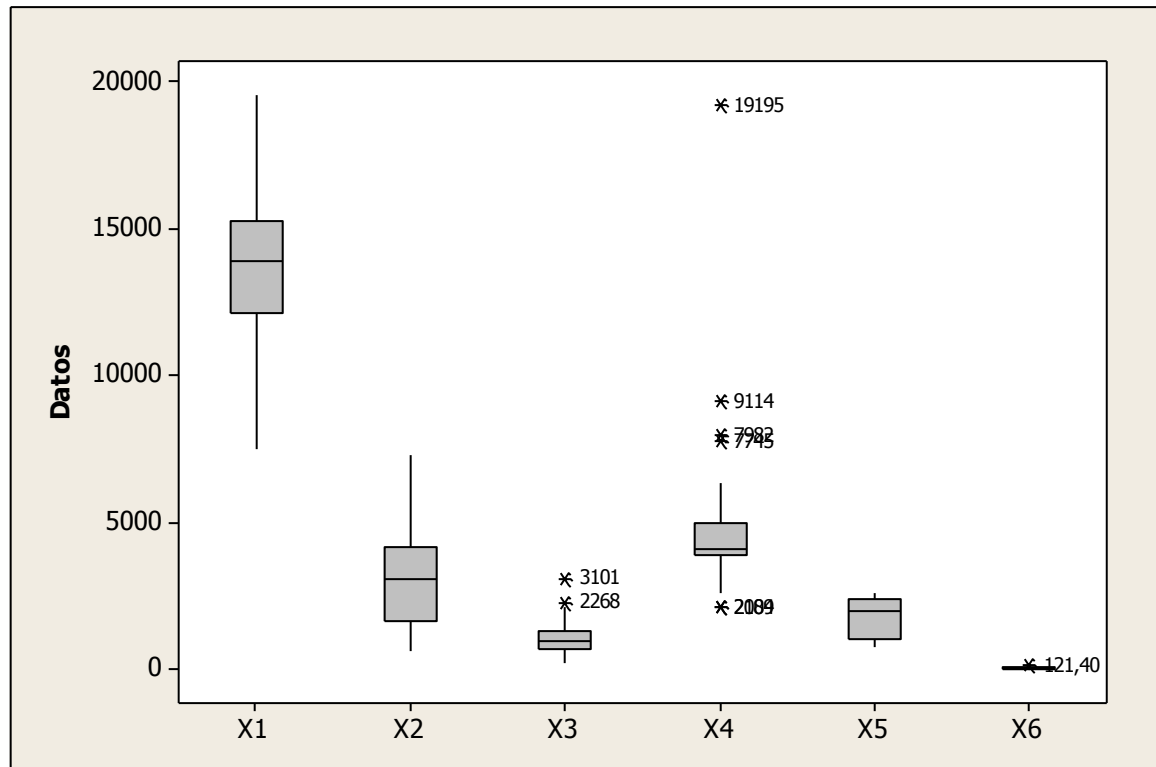
Gráfica 7: Box-plot de la variable endógena



Elaborado por: Laura Rochina

Como se puede observar la caja (Box-plot) de la variable endógena (variable producción) no se encuentran valores extremos, es decir, no tenemos la presencia de datos atípicos, además los bigotes casi simétricos indican una distribución cercana a la normal.

Gráfica 8: Box-plot de las variables exógenas



Elaborado por: Laura Rochina

Al observar la gráfica 8: encontramos que la mediana es muy distinta entre las variables exógenas, y cada una de estas variables tienen un sesgo positivo lo que nos indica una distribución asimétrica positiva, y con ello afirmamos lo interpretado en los resultados de la tabla 3. Además se observa que el conjunto de variables exógenas presenta 9 valores extremos que podríamos considerarlos como atípicos, los mismos pueden alterar los resultados obtenidos posteriormente.

Luego de haber realizado el análisis numérico descriptivo, en el que las variables exógenas presentan asimetría y tienen una dispersión muy amplia con respecto a la media de cada una de ellas, así mismo, al realizar la gráfica box.plot múltiple en donde se ha encontrado posibles datos atípicos. Se plantea una transformación de las variables con el propósito de evitar la separación de valores potenciales de la matriz de datos y estabilizar los mismos. Para ello se ha utilizado \sqrt{X} presentado en la siguiente tabla.

Tabla 4: Transformación de variables

Transformación de variables	
Variable	Transformación
Producción	Y_t
Pirola plástica	$\sqrt{X_6}$
Costo de materia prima	$\sqrt{X_4}$
Materia prima	$\sqrt{X_1}$
Materia prima perdida	$\sqrt{X_2}$
Mano de obra	$\sqrt{X_5}$
Costo de insumos	$\sqrt{X_3}$

Elaborado por: Laura Rochina

Con la nueva matriz de datos (Anexo) se realiza nuevamente un análisis descriptivo de los mismos teniendo como resultado:

Tabla 5: Estadísticos descriptivos de la matriz transformada

	Coef. Simetría	Mediana	Media	Coef.variación
Producción	-0,0002	78	79	9,80%
materia prima	-0,145	118	117	12,22%
materia prima perdida	-0,203	56	53	7,45%
costo de insumos	0,255	31	31	11,57%
costo de materia prima	2,815	64	68	3,70%
Mano de obra	-0,25	44	40	6,80%
piola plástica	0,98	5	6	10,51%

Fuente: Archivos de la Empresa Intercomunal Salinas

Elaborado por: Laura Rochina

Se puede observar en la tabla 5, los valores del coeficiente de variación son menores que el 20% y nos indican que estas variables son homogéneas.

Analizando la mediana y media de cada variable se observó que no existe diferencia significativa entonces descartamos la posible existencia de datos atípicos

La producción media mensual de hilo durante los últimos 4 años ha sido de $79^2 = 6.241$ libras. Lo que indica que no alcanzó la capacidad total de la

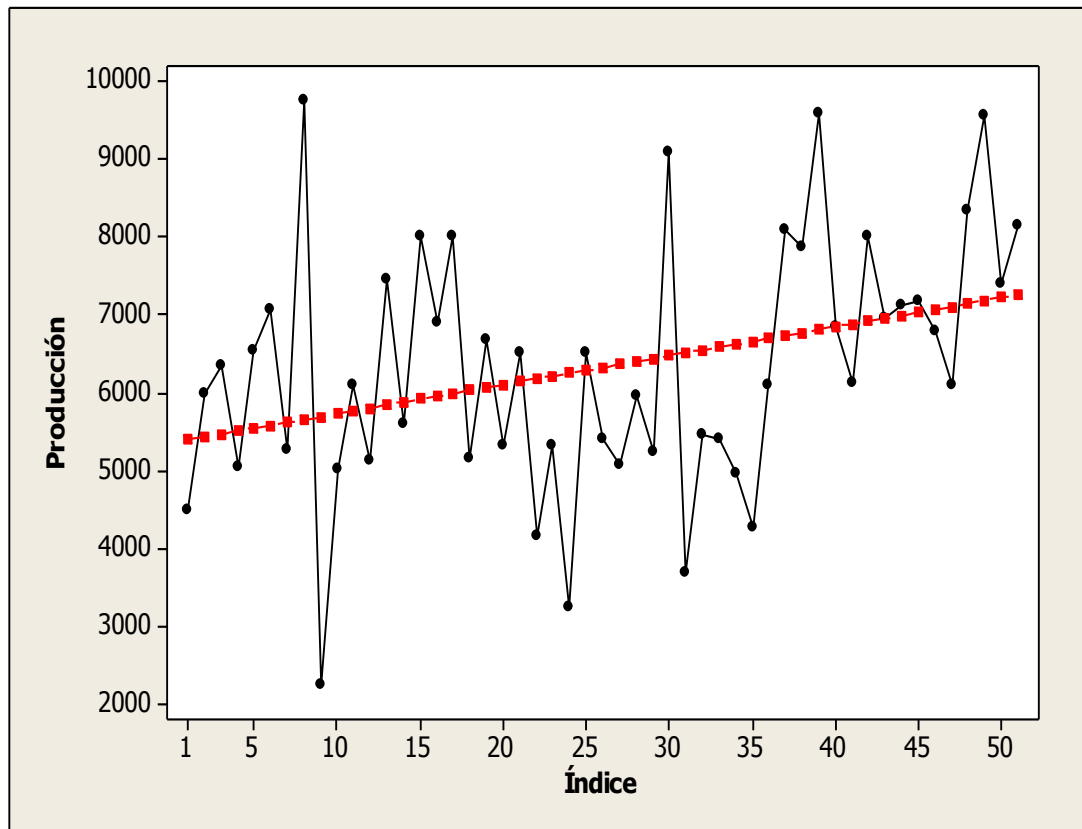
empresa que se estima en un **valor de 12.000 libras mensuales lo que representa el 52% de la producción supuesta.**

Así también la media mensual de materia prima elaborada ha sido de $117^2 = 13.689$ libras, además el promedio perdido de materia prima mensual es de $53^2 = 2.089$ libras, lo que nos indica que no existe un control en los diferentes procesos de producción, considerando que el quintal tiene 100 libras, las pérdidas mensuales se aproximan a 20 quintales de materia prima. El promedio mensual de costo de insumo para la elaboración de hilo en la empresa es de $31^2 = 961$ dólares. En promedio mensual, la planta de producción de Hilandería gasta en mano de obra un total de $40^2 = 1.600$ dólares.

3.3. Análisis de series de tiempo

Concluido AED (Análisis exploratorio de datos), utilizando algunos estadísticos descriptivos y transformado a las variables, se toma en cuenta que los datos son registrados mensualmente y se realizó el análisis de los patrones de series de tiempo, de la variable endógena.

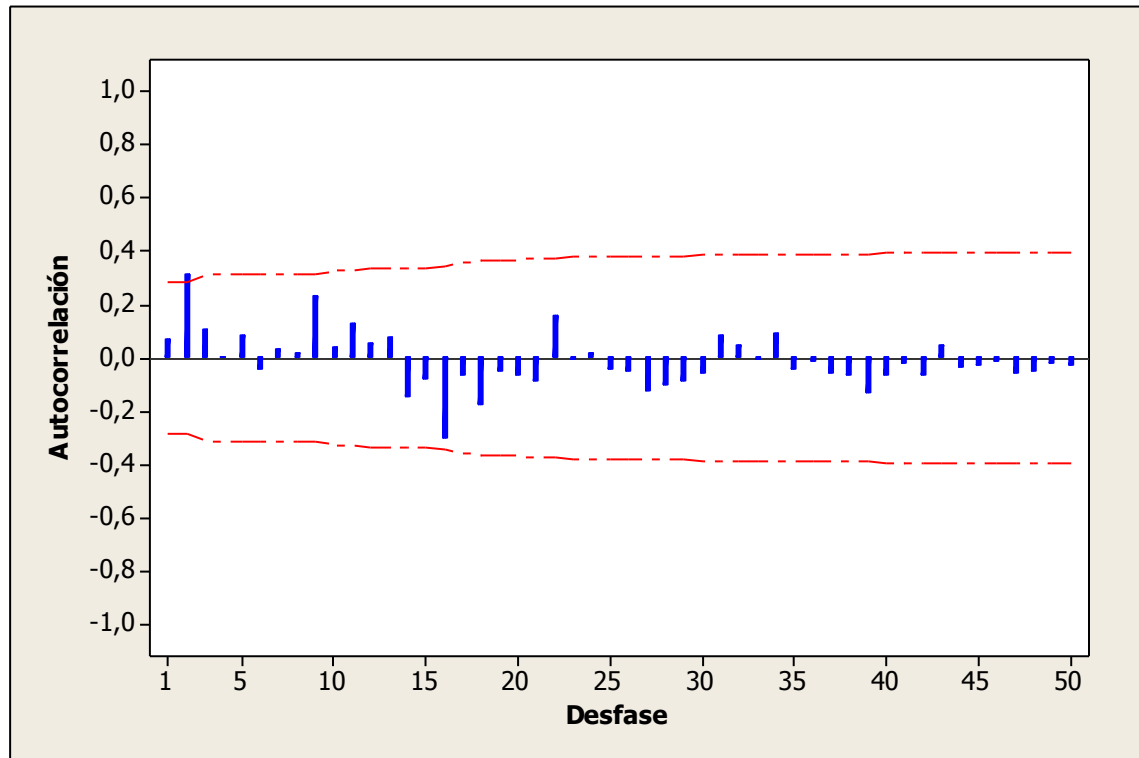
Gráfica 9: Representación de series temporales de la Producción de hilo



Elaborado por: Laura Rochina.

Como se puede observar la gráfica 9, de la variable endógena presenta tendencia.

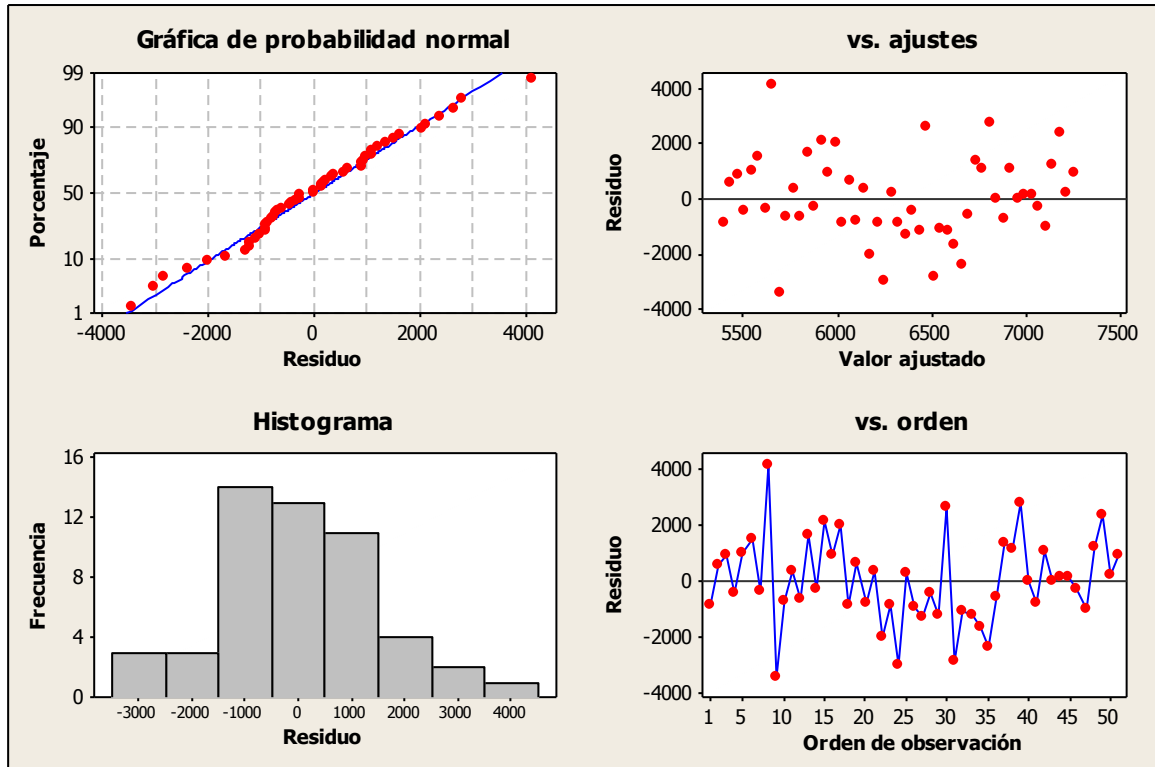
Gráfica10: Función de autocorrelación para la variable endógena (producción)



Elaborado por: Laura Rochina

Como se puede observar el correlograma de la variable endógena nos indica una serie no estacionaria, además las autocorrelaciones de muestra permanecen ligeramente grandes durante algunos periodos lo que nos indica estacionalidad con variaciones.

Gráfica 11: Representación del análisis de residuos de la variable Producción



Elaborado por: Laura Rochina

- Al observar la gráfica 11, del análisis de residuos en la variable producción se puede determinar, la normalidad de los mismos, la varianza constante y la presencia de autocorrelación.

Entonces para comprobar la presencia de autocorrelación aplicamos el estadístico Durbin Watson (d) con los parámetros:

$$n=51$$

$$k=1$$

$$\alpha=0,05.$$

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$
$$d = \frac{115261884}{114452830} = 0,992$$

$$d_L = 1,5$$

$$d_U = 1,59$$

- Donde el estadístico es menor que el límite inferior y superior, 1,5,1,59 respectivamente, entonces la producción de hilo está relacionada linealmente con los valores de la producción en otro periodo y con ello afirmamos la presencia de autocorrelación.

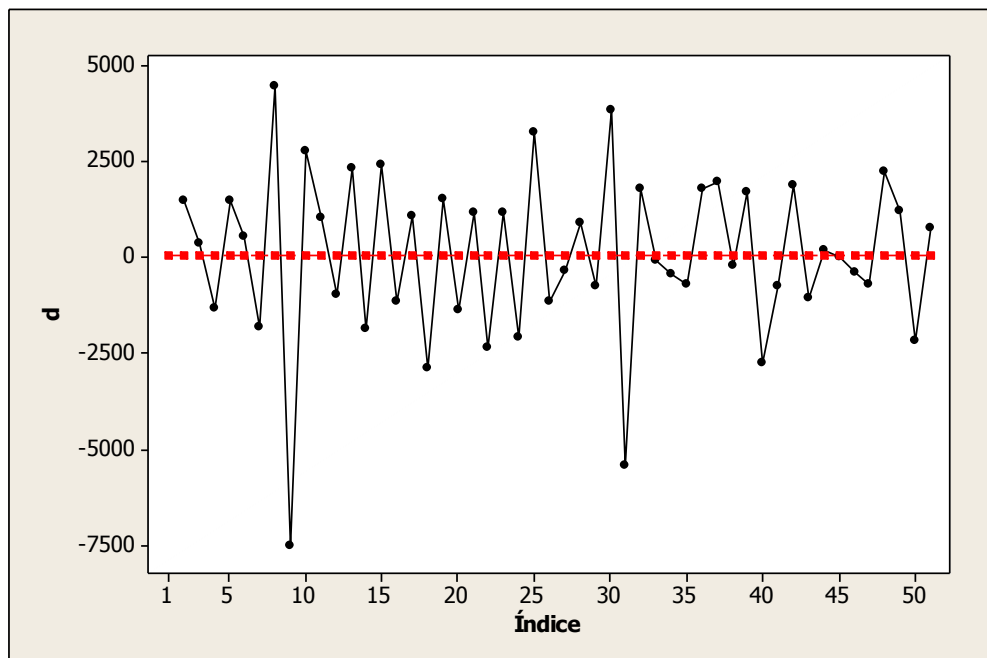
Al observar la presencia de autocorrelación positiva en la variable endógena, la normalidad en los residuos y la varianza constante, entonces para el pronóstico de la producción de hilo se plantea un modelo ARIMA.

3.3.1. Análisis de estacionariedad.

El primer paso en la identificación del modelo fue determinar si la serie es estacionaria entonces se analizó el gráfico 9, en el que claramente se observa que la serie no es estacionaria y tiene tendencia.

Entonces si Y_t =producción en el tiempo t es una serie estacionaria al tomar sus diferencias $\Delta Y_t = Y_t - Y_{t-1}$ se eliminó la tendencia como podemos observar en el siguiente gráfico.

Gráfica 12: Diferencia de la variable endógena

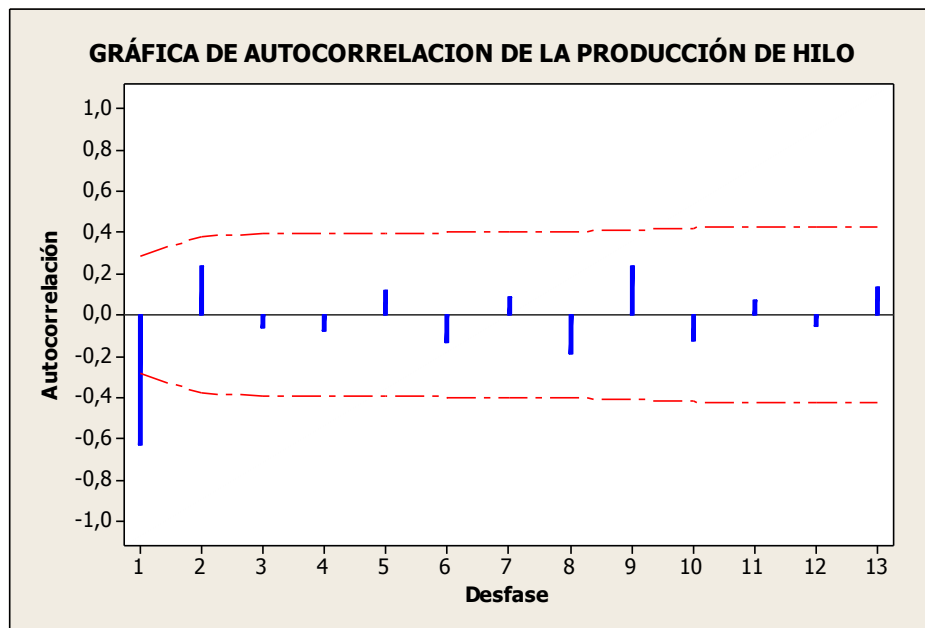


Elaborado por: Laura Rochina

3.3.2. Forma del modelo

Una vez obtenido la serie estacionaria, se identificó la forma del modelo a utilizarse para el pronóstico de la producción de hilo. Luego se definió la autocorrelación y la autocorrelación parcial, con el fin de definir los parámetros del modelo.

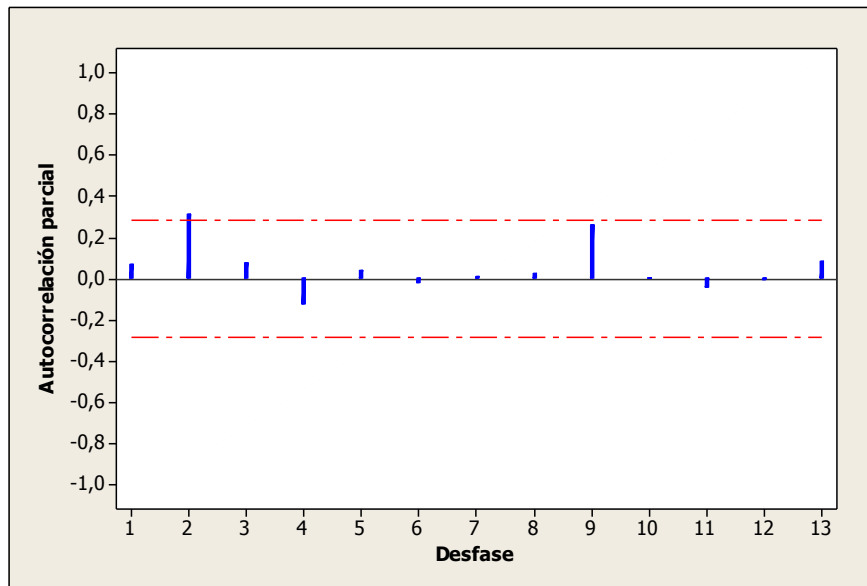
Gráfica13: Representación gráfica de Autocorrelación



Elaborado por: Laura Rochina

Al obtener la gráfica de autocorrelaciones de la producción de hilo, donde se determina el orden de la parte de media móvil. Se encuentra un desfase significativo entonces se define a $q=1$.

Gráfica 14: Representación gráfica de Autocorrelación parcial.



Elaborado por: Laura Rochina

Así también en la gráfica 14, de autocorrelaciones parciales de la producción de hilo donde se define el orden de la parte autorregresiva, se puede observar un desfase significativo y con ello $p=1$.

Entonces el modelo para el pronóstico de la producción de hilo es:

ARIMA(1,1,1)

Donde:

$p=1$ parámetro autorregresivo

$q=1$ parámetro de media móvil

$d=1$ diferencia

3.3.3. Elaboración del modelo Econométrico para la producción de la H.T.S.

Tomando el modelo ARIMA (1,1,1) para el pronóstico de la producción de hilo, se estimó los parámetros; al minimizar la suma de los cuadrados en los errores de ajuste. El modelo propuesto ARIMA (1,1,1) se ha ajustado a la serie de tiempo de la producción de hilo con 51 observaciones entonces el modelo planteado es:

$$Y_t = \frac{46,11 + 11,55Y_{t-1} - 0,8756\varepsilon_{t-1} + \varepsilon_t}{1,1155}$$

3.3.4. Evaluación del modelo para el pronóstico

Antes de usar el modelo para la realización de pronósticos de la producción de hilo, se aseguró que sea adecuado. Básicamente, un modelo es adecuado si sus residuales no pueden utilizarse para mejorar los pronósticos, es decir los residuales deben ser aleatorios.

Además se realizó el cálculo de la prueba χ^2 que se basa en los estadísticos de Ljung-Box Q que proporcionó una revisión global de la pertinencia del modelo. Esta prueba considera las dimensiones de las autocorrelaciones residuales como un grupo. La estadística de prueba Q es:

$$Q_m = n(n+2) \sum_{k=1}^m \frac{r_k^2(e)}{n-k}$$

Donde:

$r_k(e)$ = la autocorrelación residual en el retraso k.

n=4

k= 51

m=el número de retrasos de tiempo que habrán de ser evaluados

Entonces si el valor p asociado con la estadística Q es pequeño se considera que el modelo es inadecuado. Estos son los resultados obtenidos:

Tabla 6: Estadísticos para la evaluación del modelo de pronósticos

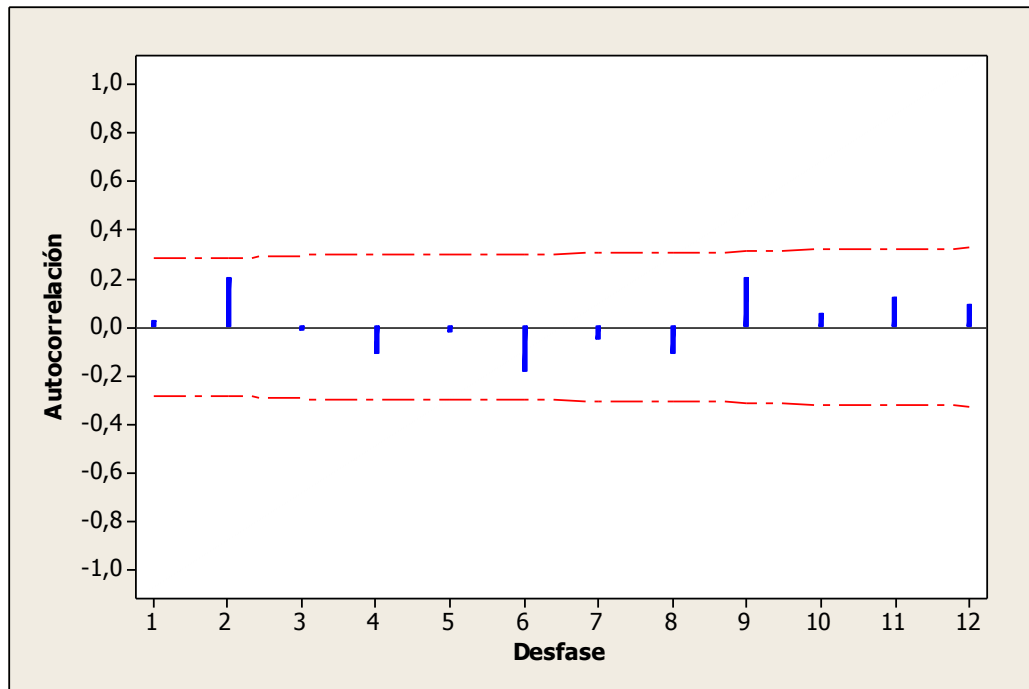
Estadística chi-cuadrada Box-Pierce(Ljung-Box)				
Desfase	12	24	36	48
Chi-cuadrada	10,1	28,9	39,1	44,1
GI	9	21	33	45
Valor p	0,345	0,116	0,216	0,508

Elaborado por: Laura Rochina

Como se puede observar en la tabla 6 de resultados el valor asociado con la estadística Q es grande $p > 0,05$, en los desfases 12, 24, 36,48 entonces se considera que el modelo propuesto ARIMA (1,1,1), para la producción de hilo es adecuado, pero se realiza el análisis de los valores p grandes. Lo expuesto

anteriormente se puede corroborar con el grafico de residuos del modelo propuesto

Gráfica 15: Representación gráfica de residuos para el modelo de producción.



Elaborado por: Laura Rochina

La gráfica 15. Muestra que no hay una autocorrelación residual significativa para el modelo propuesto, pero los repuntes sobresalientes nos indican la presencia de estacionalidad en los residuos.

Tomando en cuenta algunos desfases sobresalientes en la gráfica 15, de los residuos del modelo propuesto que nos indican la presencia de

estacionalidad, y los valores asociados del estadístico Q presentados en la tabla 5, $p > 0,05$ nos dice que el modelo es adecuado, los mismos son valores muy altos y esto se debe a la estacionalidad.

Se propone mejorar el modelo y tomar en cuenta la estacionalidad con una variación. Se plantea el método de Winters suavizamiento exponencial para variaciones de tendencia y estacionales este proporciona una forma de explicar la estacionalidad cuando los datos tienen un patrón estacional.

Entonces el modelo de pronóstico por medio del método de Winters para la producción de hilo tomando en cuenta la tendencia y la estacionalidad se desarrolla así:

Nivel estimado de serie suavizada:

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1})$$

Estimado de tendencia:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

Estimado de estacionalidad:

$$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-s}$$

Pronóstico del periodo p en el futuro:

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p}$$

Donde:

\hat{Y}_{t+p} = pronóstico para el periodo p en el futuro

L_t = nivel actual estimado

T_t = estimado de tendencia

p = número de periodo a pronosticarse

S_t = estimado de estacionalidad

s = número de estacionalidad

Tomando en cuenta: $\alpha = 0.1, \beta = 0.1, \gamma = 0.1$, mediante el criterio de mínimo error de pronósticos(MSE)

Tabla 7: Comparación de metodologías de Winters

Modelo adecuado: método de Winters		Segundo modelo: método de Winters	
Parámetros	nivel	Parámetros	nivel
Alfa(nivel)	0,1	Alfa(nivel)	0,1
Gamma(tendencia)	0,1	Gamma(tendencia)	0,1
Beta(estacional)	0,1	Beta(estacional)	0,2
Medidas de Exactitud		Medidas de Exactitud	
MAPE	6	MAPE	10
MAD	1149	MAD	1193
MSE	19,805	MSE	21,266
Tercer modelo: método de Winters		Cuarto modelo: método de Winters	
Parámetros	nivel	Parámetros	nivel
Alfa(nivel)	0,2	Alfa(nivel)	0,2
Gamma(tendencia)	0,1	Gamma(tendencia)	0,2
Beta(estacional)	0,2	Beta(estacional)	0,3
Medidas de Exactitud		Medidas de Exactitud	
MAPE	21	MAPE	22
MAD	1188	MAD	1259
MSE	21,357	MSE	22,000

Elaborado por: Laura Rochina

Al observar la tabla 7. Se elige el primer modelo por el método de Winters con un MAPE del 6% y un MSE de 19,805 ya que en comparación de los demás estos presentan los más bajos valores de las medidas de precisión del resto.

3.3.5. Utilización del modelo en los pronósticos.

La empresa Hilandería Intercomunal Salinas tiene el propósito de incrementar su producción para ello se pronosticó la producción de hilo con el método de

Winters, $\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p}$, para el periodo de un año y así puedan formular políticas para cubrir la producción programada. Entonces a partir del modelo obtenido se desarrolla el método de Winters:

Tabla 8: Valores mediante el método de Winters para la producción de hilo

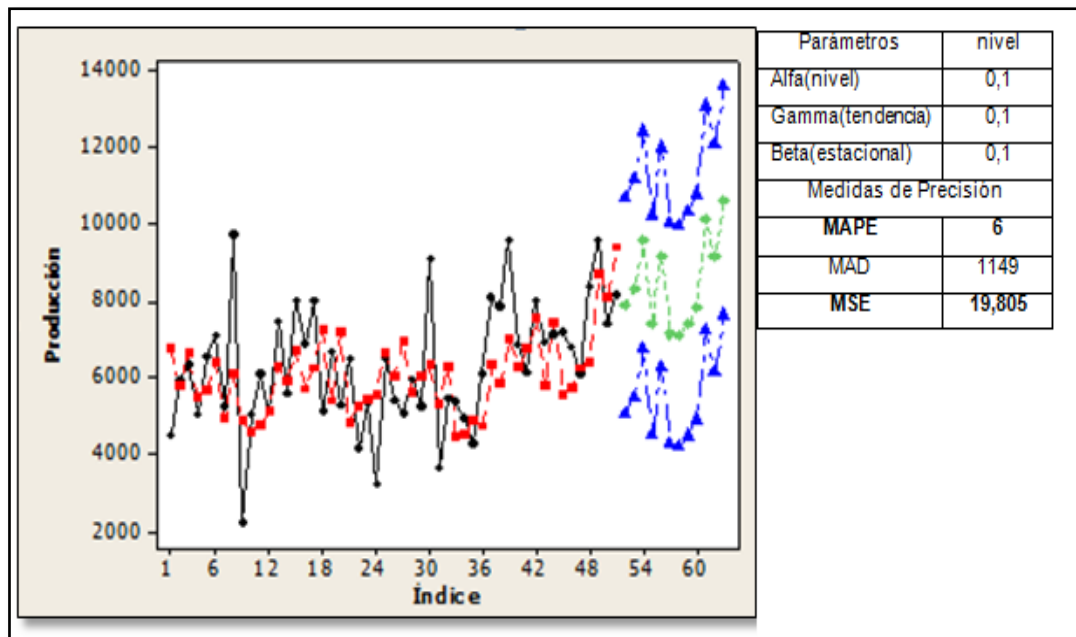
Año	t	Y_t	L_t	T_t	S_t	\hat{Y}_{t+p}	e_t
2008	1	4492	569,95	-46,32	1,11	4767,10	-275,10
	2	5980	566,58	-44,52	1,03	5795,80	184,20
	3	6352	559,19	-47,01	1,18	6646,07	-294,07
	4	5045	550,13	-51,51	0,98	5490,80	-445,80
	5	6530	553,45	-43,03	1,05	5650,79	879,21
	6	7076	554,85	-37,49	1,18	6427,58	648,42
	7	5253	554,05	-34,22	0,90	4958,93	294,07
	8	9753	584,97	-1,00	1,16	8885,38	867,62
	9	2241	552,08	-32,89	0,80	2934,51	-693,51
	10	5023	554,67	-27,34	0,84	4561,93	461,07
	11	6091	567,23	-12,15	0,89	5776,27	314,73
	12	5119	565,62	-12,00	0,90	5105,02	13,98
2009	13	7457	575,12	-1,45	1,13	6883,29	573,71
	14	5605	572,12	-4,50	1,02	5919,39	-314,39
	15	8023	582,24	6,46	1,20	7732,01	290,99
	16	6892	595,13	18,30	1,00	5728,82	1163,18
	17	8005	613,41	34,80	1,08	7271,69	733,31
	18	5143	598,00	16,68	1,15	5281,84	-138,84
	19	6687	614,62	30,57	0,92	6430,48	256,52
	20	5303	601,63	14,52	1,13	6166,70	-863,70
	21	6501	623,17	35,42	0,82	5827,19	673,81
	22	4160	614,51	22,31	0,82	4258,98	-98,98
	23	5318	614,39	20,67	0,88	5463,58	-145,58
	24	3246	591,93	-5,05	0,87	3566,53	-320,53
2010	25	6524	589,18	-6,42	1,13	6679,06	-155,06
	26	5404	582,81	-12,51	1,01	6028,28	-624,28
	27	5070	565,42	-28,30	1,17	5960,60	-890,60
	28	5967	566,34	-24,88	1,01	5624,89	342,11
	29	5241	555,97	-32,53	1,06	6064,13	-823,13
	30	9101	576,43	-8,53	1,19	8345,50	755,50
	31	3675	558,36	-26,28	0,90	4313,29	-638,29
	32	5465	548,08	-33,58	1,12	6291,99	-826,99
	33	5392	555,38	-22,69	0,84	4493,80	898,20
	34	4955	558,90	-17,77	0,83	4550,37	404,63
	35	4272	549,72	-25,11	0,87	4921,04	-649,04
	36	6098	562,18	-9,46	0,89	5740,34	357,66
2011	37	8085	576,76	6,05	1,16	7335,35	749,65
	38	7866	597,63	25,83	1,04	7858,94	7,06
	39	9580	621,19	47,90	1,20	9003,04	576,96
	40	6840	632,83	53,28	1,01	6299,81	540,19
	41	6111	631,72	47,04	1,05	6774,01	-663,01
	42	8008	639,14	50,68	1,20	7574,64	433,36
	43	6942	657,70	63,66	0,91	6777,77	164,23
	44	7130	661,44	60,97	1,12	7431,31	-301,31
	45	7170	686,18	79,65	0,86	6602,18	567,82
	46	6793	706,44	92,21	0,84	5751,91	1041,09
	47	6097	713,78	90,42	0,87	6253,12	-156,12
	48	8353	744,91	112,09	0,91	7426,74	926,26
2012	49	9556	762,86	119,18	1,17	8737,03	818,97
	50	7389	768,34	112,41	1,04	8096,29	-707,29
	51	8156	769,35	102,17	1,19	9038,63	-882,63

Elaborado por: Laura Rochina

Fuente: Archivos de la empresa

Al observar la tabla 8. Se compara los valores de la producción real con los estimados y observamos que los datos estimados mantienen el patrón estacional en el 4to mes de los años 2009,2010 con valores muy semejante, así también en el 3er mes de los años 2011 y 2012, entonces la estacionalidad presentada en los estimados es muy variada.

Gráfica 16: Resultados gráficos Método de Winters



Elaborado por: Laura Rochina

Al observar la gráfica 16 de los valores ajustados con la producción real encontramos que los pronósticos siguen un patrón con tendencia de crecimiento pero los mismos presentan puntos donde nos indican estacionalidad. Además los límites inferior y superior de los pronósticos indicados de color azul se encuentran estrechos.

Tabla9: Pronósticos de la producción de hilo

AÑO	MESES	PRONÓSTICOS	L.INFERIOR	L.SUPERIOR
2012	ABRIL	7894,2	7079,21	8709,1
2012	MAYO	8316,4	7490,67	9142,2
2012	JUNIO	9575,9	8738,71	10413,2
2012	JULIO	7389,6	6540,24	8239
2012	AGOSTO	9146,9	8284,74	10009,1
2012	SEPTIEMBRE	7142,1	6266,5	8017,8
2012	OCTUBRE	7079	6189,31	8868,8
2012	NOVIEMBRE	7416,7	6512,24	8321,2
2012	DICIEMBRE	7854,5	6934,33	9065,6
2013	ENERO	10155,1	9219,33	10090,9
2013	FEBRERO	9137,2	8137,2	10089,4
2013	MARZO	10614,6	9645,14	10584,1

Elaborado por: Laura Rochina

Entonces se dijo que en el mes de Abril 2012 la Hilandería Intercomunal Salinas producirá alrededor de 7.894,2 libras de hilo en el mes de Abril, en el mes de mayo del mismo año 8.316,4 libras, y consecuentemente para los meses posteriores hasta llegar a los meses, Febrero con 9.137,2 libras y Marzo, del 2013 con una producción de 10.614,6 libras. Teniendo como media mensual de producción 9.226,85 libras, durante un año.

3.4. Planteamiento del análisis estructural

3.4.1. Análisis de correlación.

Esta técnica estadística nos ayudará a comprobar la existencia o no de la relación entre dos o más variables que implica la producción de hilo, y encontrar la mejor serie posible de variables exógenas.

Tabla 10: Correlación de las variables endógena y exógenas

	Y	X1	X2	X3	X4	X5	X6
Y	1						
X1	0,64	1					
X2	0,11	0,38	1				
X3	0,59	-0,17	0,13	1			
X4	0,34	-0,06	0,01	0,19	1		
X5	0,34	0,01	0	0,2	0,58	1	
X6	0,32	0,31	0,11	0,13	-0,02	0,63	1

Elaborado por: Laura Rochina

Como podemos observar en la tabla 10. Las autocorrelaciones de las variables en estudio, revela que las variables X_1 (Materia prima) y X_3 (Costo de insumos) tienen una relación más cercana con la producción de hilo, seguido de X_5

(mano de obra), X_4 (costo de materia prima), X_6 (piola plástica) y X_2 (materia prima perdida).

También podemos observar que piola plástica tiene una relación de 0,636 con mano de obra, entonces decimos que existe colinealidad moderada entre las variables mencionadas.

3.4.2. Análisis de regresión

Para la hilandería Intercomunal Salinas se realizó la propuesta del análisis regresión (múltiple) que basa en las variables exógenas en función del tiempo, con coeficientes que pueden cambiarse a través del tiempo, para el mismo se utilizó la nueva matriz de datos transformados (Anexo 1), en el cual la tendencia y estacionalidad presentada en algunas variables son modelados explícitamente.

Entonces para el análisis propuesto se presenta el siguiente modelo general de regresión múltiple:

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 \\ Y &= -5,157 + 0,334 X_1 - 0,135 X_2 + 0,953 X_3 + 0,203 X_4 + 0,467 X_5 + 0,375 X_6 \end{aligned}$$

Donde,

Y = Producción de hilo

X_1 = Materia prima

X_2 = Materia prima perdida

X_3 = Costo de insumos

X_4 = Costo de materia prima

X_5 = Mano de obra

X_6 = Piola plástica

Tabla 11: Tabla ANOVA

ANOVA					
	Df	SS	MS	F	Significancia F
Regresión	6	59664072	9944011,9	6,16549945	9,44E-05
Residual	44	70965301	1612847,8		
Total	50	130629373			
R ajustado	0,45				
Error estándar	1.342				

Elaborado por: Laura Rochina

El valor F (6,16) se usó para probar la significancia de regresión. El valor calculado F y su pequeño valor p (9,44331E-05) muestra que la regresión es significativa, entonces la producción de hilo se relaciona al menos con alguna de las variables independientes.

Así también el coeficiente de determinación $R^2 = 0.45$, nos quiere decir que alrededor del 45% de la variación producción de hilo, se explica por el modelo planteado. Lo cual no es satisfactorio.

Tabla 12: Resultados por variables exógenas

	Coeficientes	t Stat	P-value	VIF
Intercepto	-5,157	0,436	0,665	
Materia prima	0,334	4,142	0	1,59
Materia prima perdida	-0,136	-1,12	0,269	1,24
Costo de insumos	0,954	3,066	0	1,16
Costo de materia prima	0,203	-1,314	0,196	1,76
Mano de obra	0,467	0,971	0,337	3,03
Piola plástica	0,375	0,429	0,67	2,38

Elaborado por: Laura Rochina

Como se puede observar en los resultados del modelo general, las variables materia prima y costo de insumos son significativos, mientras que las demás variables podrían descartarse del modelo.

3.4.2.1. Estudio de multicolinealidad

Para el análisis de la relación lineal entre dos o más variables se utilizó el factor de expansión de la varianza(VIF) resultados presentados en la tabla 8. Entonces encunto a las variables mano de obra con un VIF=3,03 y piola

plástica con $VIF=2,38$ decimos que existe una información redundante, es decir están relacionadas linealmente con las variables explicativas restantes. Cabe recalcar aquellos resultados se pudo observar en el análisis de correlación anteriormente.

Con el fin de definir un modelo donde nos explique una proporción adecuada y no presente problemas de multicolinealidad, además considerando variables que tienen una relación alta con la producción de hilo se propone la siguiente regresión:

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_4 X_4 + \beta_6 X_6 \\ Y &= 13,26 + 0,29 X_1 + 0,42 X_3 + 0,38 X_4 + 0,47 X_6 \end{aligned}$$

Donde,

Y = Producción de hilo

X_1 = Materia prima

X_3 = Costos de insumos

X_4 = Costo de materia prima

X_6 = Piola plástica

3.4.3. Significancia del modelo

Tabla 13: ANOVA

ANOVA					
	Df	SS	MS	F	Significancia F
Regresión	2	43857009	2E+07	12,13	5,45E-05
Residual	48	86772364	2E+06		
Total	50	1,31E+08			
R-ajustado	0,68				
Error estándar	1,346				

Elaborado por: Laura Rochina

Se encuentra un valor $F(12,130)$ que se usó para probar la significancia del nuevo modelo de regresión propuesto con las variables; materia prima y costos de insumos . El valor calculado F y su pequeño valor p ($5,44721E-05$) muestra que la regresión es significativa, entonces la producción de hilo se relaciona al menos con alguna de las variables independientes.

Así también el coeficiente de determinación en este caso es de $R^2 = 0.68$, nos quiere decir que alrededor del 68% de la variación producción de hilo, se explica por el modelo planteado.

Al realizar el análisis del error estándar de la estimación ya que es una medida de dispersión de los valores de Y respecto a la función de regresión ajustada.

$$S_{y.x's} = 1,344$$

Esto nos indica la diferencia entre la producción de hilo real y los ajustados, son considerablemente pequeños.

3.4.4. Variables explicativas individuales:

En el modelo propuesto, el coeficiente de una variable endógena mide el efecto parcial, o neto, de esa variable exógena en la respuesta de la variable endógena, manteniendo a las otras variables exógenas constantes.

Como el modelo nuevo planteado fue considerado significativo, es de interés examinar la significancia de las variables explicativas. Para el mismo se analiza un valor p apropiado.

Tabla 14: Tabla de estadísticos para la significancia del modelo.

Variable	Coeficientes	t Stat	P-value	VIF	Significancia
Intercepto	13,26	0,151	0,88		No significativo
Materia prima	0,29	4,229	0,0001	1,031	Significativo
costo de insumos	0,422	3,249	0,002	1,031	Significativo
Costo de materia prima	0,38	7,18	3,13E-09	1,082	Significativo
Pirola plástica	0,47	5,2	3,71E-06	1,043	Significativo

Elaborado por: Laura Rochina

Al observar la tabla 14; encontramos valores $p < 0,05$ en las variables exógenas propuestas, estos nos indica que significativamente son distintos de cero. Por ello no deben excluirse del modelo. Finalmente el $VIF = 1,031$ común para las dos variables explicativas en el segundo modelo indica que la multicolinealidad ya no es un problema

La interpretación de $\hat{\beta}_1 = 0,29$ sería que por el incremento de una libra de materia prima, cuando los costos se mantenga constantes, la producción promedio de hilo aumenta en 2,9 libras. Del mismo modo, el valor $\hat{\beta}_3$ de 0.4223 significa que si los costos de insumos se incrementan en 10 dólares cuando la cantidad de materia prima se mantenga constante, el costo de producción de hilo se incrementará en un promedio de 4 dólares con 20ctv por libra.

CAPÍTULO

IV

4. CONCLUSIONES Y RECOMENDACIONES

4.1. CONCLUSIONES

- La hilandería Intercomunal Salinas perteneciente al Cantón Guaranda, carecía de estudios econométricos y estadísticas, de la producción de hilo desde su creación, por lo que se planteó una investigación orientada a pronosticar la producción y realizar un análisis de las variables para tomar decisiones en beneficio de la empresa.
- Del análisis descriptivo de las variables se ha encontrado que la producción media mensual de hilo durante los últimos 4 años ha sido de 6.241 libras. Esto indica que no alcanzó la capacidad total de la empresa que se estima en un **valor de 12.000 libras mensuales lo que representa el 52% de la producción supuesta.**
- Además el promedio perdido de materia prima mensual es de 2.089 libras, lo que nos indica que no existe un control en los diferentes procesos de producción, considerando que el quintal tiene 100 libras, **las pérdidas mensuales se aproximan a 20 quintales de materia prima.**

- Se ha encontrado el modelo econométrico para el pronóstico de la producción de hilo mediante el método de Winters con $\alpha=0,1$, $\beta=0,1$ y $\gamma=0,1$ que toma en cuenta la tendencia y la estacionalidad dados por:

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p}$$

- El pronóstico de producción mensual se realizó para 12 meses, teniendo como resultado una producción 7.894,2 libras de hilo en el mes de Abril, en el mes de mayo del mismo año 8.316,4 libras, y consecuentemente para los meses posteriores hasta llegar a los meses, Febrero con 9.137,2 libras y Marzo, del 2013 con una producción de 10.614,6 libras. Teniendo como media mensual de producción 9.226,85 libras, durante un año.
- Del análisis estructural de las variables en estudio se determinó un modelo de regresión múltiple $Y = 0,29X_1 + 0,42X_3 + 0,38X_4 + 0,47X_6$ que determinó la significancia de las variables en relación con la producción de hilo y estas corresponden a las variables X_1 = materia prima, X_3 = costo de insumo, X_4 = costo de materia prima y X_6 = piola plástica.

4.2. RECOMENDACIONES.

- La empresa debe planificar la producción estimada de 9.227 libras mensuales encontrada en esta investigación por el método de Winters que se ha realizado en base a la producción de años anteriores más no las 12.000 libras que indica en los manuales de las máquinas de la empresa habiendo una subutilización de la producción.
- Realizar un mayor control en los diferentes procesos de producción para evitar las pérdidas de materia prima ya que las mismas son de 20 quintales mensuales, y estos muy elevados.
- En el análisis estructural, no se tomaron en cuenta las variables X_2 = materia prima perdida, X_5 =mano de obra, se eliminaron por no ser significativos. Se recomienda estudiar otras variables que se relacionen con la producción de hilo.
- Se recomienda fortalecer las investigaciones, sobre todo estadísticas como la técnica Box Jenkins o ARIMA, que ayudan en la planificación de la producción de hilo en la empresa.

- Para una mejor producción de hilo en H.T.S. y mejoramiento del mismo se recomienda estudios de control de calidad.

BIBLIOGRAFÍA

BIBLIOGRAFIA

1. **-ALLEN, L.**, Estadística Aplicada de los Negocios y la Economía., 2da ed., Madrid-España., McGraw-Hill., 2000., Pp. 30-74.
2. **-CANAVOS, G.**, Análisis de Regresión., 4ta ed., México- D.F. México., McGraw-Hill., 1999., Pp. 25-43.
3. **-CONGACHA J.**, ESTADÍSTICA APLICADA A LA EDUCACIÓN., 1era., Madrid-España., LAP LAMBERT., 2012., Pp.30-79.
4. **-DAMODAR N, GUJARATI, DAWN C.**, Econometría. 5ta ed., Madrid-España., McGraw-Hill., 2010., Pp. 20-98.
5. **-GLYN G.**, Matemática Avanzada para Ingeniería., 2da ed., México-D.F. México., Pearson Educación., 2002., Pp. 43-51.
6. **-JOHN E. HANKE.**, Pronósticos en los Negocios., 9na ed., México-D.F. México., Pearson Educación., 2008., Pp. 20-307.
7. **-MONTGOMERY, Douglas C, RUNGER G.**, Probabilidad y Estadística Aplicada a la Ingeniería., 2ª ed., Madrid-España., McGraw-Hill., 2003., Pp. 90-100.

8. **-PINDYCK, R.**, Econometría, Modelos y Pronósticos., 4ta ed., México-D.F.

México., McGraw-Hill., 2001., Pp.103-204.

9. **-PULIDO A. PERZ.J.**, Modelos Econométricos con Eviews., 5ta ed., México-

DF. México., Pearson Educación., 2005.

10. **-WEBSTER L.**, Estadística Aplicada a los Negocios y la Economía., 3ra ed.,

Bogotá-Colombia., Camargo., 2002., Pp.33-41

BIBLIOGRAFÍA DE INTERNET

11.-METODOLOGÍA DE INVESTIGACIÓN

http://www.upec.edu.ec/citte/archivos/manuales/MANUAL_ELABORACION_TESIS.pdf

2012/04/03.

12.-ESTADÍSTICOS DESCRIPTIVOS

http://www.fca.unam.mx/docs/apuntes_matematicas/34.%20Estadistica20%Descriptiva.pdf

2012/05/15.

13. -ANÁLISIS EXPLORATORIOS DE DATOS

<http://www.fisterra.com/mbe/investiga/10descriptiva/10descriptiva.asp>

2012/08/05.

14. -ANÁLISIS ESTADÍSTICO UNIDIMENSIONAL

<http://www.tuveras.com/estadistica/estadistica02.htm>

2012/08/13.

15. -SERIES DE TIEMPO

<http://www.dii.uchile.cl/~ris/RISXXIII/Buschiaz089.pdfseries de tiempo>

2012/09/05.

16. -MODELOS ECONOMETRICOS

<http://www.econometricos.com.ar/wpcontent/uploads/2007/12/Espacio-de-Estado-2004.pdf>

2012/10/12.

17.-MODELOS ECONÓMICOS

http://www.secmca.org/INVESTIGACIONES_ECONOMICAS/InvestigacionesSECMCA/ModeloEconometricoCrecimientoEconomicoInflacionCARD.pdf

11/10/2012.

18. -COMPONENTES DE UN MODELO ECONOMÉTRICO

<http://www.fca.uach.mx/Documentos/Guias/modeloseconometricos.pdf>

2012/10/13.

19. -MODELO ARIMA

http://www.ccee.edu.uy/ensenian/catectr/material/arima_1.pdf

2012/11/16.

20. -MÉTODO DE WINTERS

http://www.ub.edu/aplica_infor/spss/cap8-5.htm

2013/01/05.

21.-PRODUCCIÓN DE HILO

<http://www.redtextilargentina.com.ar/index.php/hilos/h-produccion>

2013/01/20.

22. -OPTIMIZACIÓN DE RECURSOS

<http://www.itecam.edu.mx/principal/sylabus/fpdb/recursos/r5300.PDF>

2013/02/23.

23. -ORGANIZACIONES DE LA PARROQUIA SALINAS

<http://www.agriculturesnetwork.org/magazines/latin-america/mas-que-el-dinero/la-experiencia-de-desarrollo-de-salinas-de>

2013/02/26.

ANEXOS

Anexo 1

MATRIZ DE DATOS TRANSFORMADOS

Producción	materia prima	materia prima	costo de insumos	costo de materia	Mano de obra	piola plastica
67,02	104,88	31,38	31,27	45,92	27,93	4,36
77,33	115,90	35,83	18,03	62,49	28,71	4,35
79,70	118,57	60,01	26,21	79,70	28,84	4,53
71,03	121,98	64,79	18,00	61,57	29,83	4,79
80,81	122,87	58,07	27,93	55,86	28,81	4,39
84,12	131,93	70,87	44,20	55,90	27,93	4,11
72,48	120,96	59,34	33,14	54,23	28,30	4,37
98,76	136,02	77,03	29,95	60,68	29,83	4,51
47,34	103,89	44,43	21,28	50,73	30,08	4,13
70,87	128,46	58,48	15,20	54,91	31,14	4,13
78,04	117,74	63,81	31,37	45,65	29,44	4,36
71,55	110,09	33,96	25,53	62,97	30,98	4,31
86,35	111,11	42,40	33,32	62,46	40,96	6,31
74,87	110,20	63,53	36,29	69,86	32,30	5,30
89,57	121,00	55,56	45,85	65,80	35,13	5,57
83,02	112,10	75,64	31,14	63,88	33,23	5,00
89,47	127,78	51,57	29,07	63,51	32,83	4,36
71,71	126,44	25,14	18,03	60,84	39,70	6,71
81,77	116,82	52,50	34,68	70,46	36,95	5,30
72,82	121,38	60,34	31,10	62,54	35,75	4,84
80,63	107,97	34,80	18,00	64,19	36,11	4,35
64,50	110,95	52,35	32,95	64,23	35,83	4,40
72,92	105,30	27,04	31,38	62,55	36,18	4,50
56,97	112,69	42,45	36,33	63,80	34,50	4,59
80,77	123,57	53,87	34,81	65,27	45,33	10,70
73,51	115,87	59,76	44,56	69,22	43,31	7,78
71,20	124,90	75,47	35,20	60,84	44,38	7,10
77,25	107,42	85,24	33,29	65,73	44,74	4,57
72,39	118,63	62,63	28,04	66,90	45,71	8,73
95,40	139,60	25,81	25,98	76,82	44,42	7,09
60,62	95,55	40,45	22,47	76,42	44,92	4,54
73,93	104,08	56,99	36,07	89,34	44,58	4,25
73,43	95,42	49,50	33,38	62,23	46,13	4,37
70,39	86,70	51,42	45,88	88,01	44,47	4,29
65,36	92,21	30,95	18,84	73,22	45,71	4,24
78,09	111,98	45,80	32,16	62,97	44,58	6,34
89,92	105,74	39,09	36,23	64,25	48,56	11,02
88,69	118,33	72,48	47,63	69,43	48,95	9,38

97,88	121,13	68,18	34,43	138,55	50,70	3,92
82,70	123,17	63,94	29,21	65,70	49,01	9,49
78,17	138,20	70,84	31,70	73,86	50,81	8,87
89,49	114,11	25,44	41,71	95,47	50,81	4,93
83,32	139,84	69,98	26,45	64,07	49,32	8,34
84,44	139,76	66,61	16,35	74,06	49,32	6,16
84,68	110,36	54,94	29,72	62,64	49,32	8,10
82,42	131,38	64,28	18,71	63,95	48,39	7,80
78,08	124,05	71,87	26,15	77,18	49,32	8,07
91,39	118,44	30,73	14,42	69,96	49,32	7,72
97,75	110,65	32,36	55,69	77,07	49,90	6,68
85,96	117,88	45,32	45,55	63,76	48,86	7,73
90,31	120,55	56,65	44,56	69,97	49,87	6,30